Technical Report 09-001

Interfacing Network Simulations and Empirical Data

2LT Joshua Lospinoso, MAJ Ian McCulloh,
MAJ Anthony Johnson

U.S. Military Academy, West Point NY

May 2009

United States Military Academy
Network Science Center

U.S. Military Academy
Network Science Center

Authorized and approved for distribution:

DANIEL EVANS
Deputy Director

LTC JOHN GRAHAM, Ph.D.
Director

Technical review by

Craig Lennon, Ph.D., Department of Mathematical Sciences, U.S. Military Academy
MAJ John Roginski, Department of Mathematical Sciences, U.S. Military Academy

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by the U.S. Military Academy Network Science Center. Please address correspondence concerning distribution of reports to: Network Science Center, U.S. Military Academy, 646 Swift Road, West Point, NY 10996

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Military Academy Network Science Center.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
25 May 2009	Technical Report	June 2008 – May 2009

4. TITLE AND SUBTITLE
Interfacing Network Simulations and Empirical Data

5a. CONTRACT NUMBER n/a
5b. GRANT NUMBER n/a
5c. PROGRAM ELEMENT NUMBER n/a

6. AUTHOR(S)
Joshua Lospinoso, Ian McCulloh, Anthony Johnson

5d. PROJECT NUMBER IKE03-02
5e. TASK NUMBER n/a
5f. WORK UNIT NUMBER n/a

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Mathematical Sciences,
U.S. Military Academy

8. PERFORMING ORGANIZATION REPORT NUMBER
n/a

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USMA Network Science Center

10. SPONSOR/MONITOR'S ACRONYM(S)
USMA NSC

11. SPONSOR/MONITOR'S REPORT NUMBER(S) 09-001

12. DISTRIBUTION / AVAILABILITY STATEMENT
Unlimited Distribution

13. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT
This project tests Social Network Models for longitudinal data against empirical data using an original statistical test to determine the effectiveness of various models at reproducing networks. The Link Probability Model (LPM) is introduced as a viable model for the reproduction of social networks in dynamic equilibrium. We survey social network simulation packages and find that Construct uses a continually updated LPM as its stochastic engine, further establishing the LPM's viability as a social network model. We use actor oriented models to estimate statistically significant behavior on empirical networks and provide guidance on future extensions into multi-agent simulation, which is a rapidly growing area of research. Our findings rely on various empirical datasets and provide analytical results on the nature and structure of the social networks observed in them.

15. SUBJECT TERMS
Social Network Analysis, Actor Oriented Social Networks, Social Network Simulations, Command and Control, ELICIT, IkeNET

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Tish Torgerson
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UL	44	**19b. TELEPHONE NUMBER** *(include area code)*
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED			845-938-8238

Technical Report 09-001

Interfacing Network Simulations and Empirical Data

2LT Joshua Lospinoso, MAJ Ian McCulloh, MAJ Anthony Johnson

U.S. Military Academy, West Point NY

U.S. Military Academy Network Science Center
646 Swift Road, West Point, NY 10996

25 May 2009

ACKNOWLEDGEMENT

The author would like to thank Dr. Kathleen Carley and the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University for their support on this project as well as Dr. Tom Snijders and the Department of Statistics at the University of Groningen.

INTERFACING NETWORK SIMULATIONS WITH EMPIRICAL DATA

EXECUTIVE SUMMARY

Social network analysis (SNA) is the mathematical methodology of quantifying connections between individuals and groups. It has become an important analytic tool for analyzing terrorist networks, friendly command and control structures, arms trade, biological warfare, and the spread of diseases, among other applications. This analysis provides a wealth of information about how individuals in a network interact with each other. Much of the power of SNA is derived from our ability to make prescriptions and predictions about network behavior. There are advanced simulation packages readily available to conduct this analysis, but it is particularly difficult to validate these simulation models. It is desirable to model the actor behavior from the simulation in a statistical context and estimate relevant parameters from empirical data. In this way, simulations could be grounded in robust analysis of real world data.

We have developed and surveyed a number of statistical frameworks including the Link Probability Model (LPM), the Exponential Random Graph Model (ERGM), and the Actor Oriented Model (AOM). Each of these models has parameters that can be empirically obtained from social network data to advise accurate simulations. To facilitate our analysis, we created statistical tests and empirical frameworks that contribute to future researchers' abilities to conduct comparison studies.

Procedure

This project utilized data collected from the IkeNet (McCulloh et al, 2008) and ELICIT (Lospinoso et al, 2009) experiments conducted at the United States Military Academy, as well as many popular data sets from the SNA literature. We construct a simple, baseline statistical model called the LPM as well as a robust statistical test to determine how well simulated network data fits empirically observed data. We compare the LPM to the ERGM using various data sets. We then utilize an AOM specification to empirically estimate rate functions that can be used to advise proper model specification within multi-agent simulation, then provide future directions for studying the interface of AOM and constructuralist-based simulation packages like Construct.

Findings

This report finds that while LPMs perform better than ERGMs in many of the data sets we encountered across multiple domains, the AOM has the potential to outperform both. Future work will be needed to test the efficacy of AOM in providing robust estimates of behavioral parameters for use in accurate multi agent simulations. We also reinforce the literature's finding that the AOM is able to determine statistically significant sociological phenomena within a particular dataset, as well as bridge the gap between empirically estimated parameters from a social network model into a workable simulation package.

Utilization and Dissemination of Findings

This research is an emerging area of Network Science. As such it has been and will continue to be presented at a variety of academic conferences. The following is a list of conference proceedings and publications which directly contributed to this report, and represent dissemination of its content in conference settings:

1. Lospinoso, J. "Constrained Communication Patterns: An Empirical Estimation of Actor Oriented Social Network Behavior." International Conference on Information & Knowledge Engineering Proceedings, Las Vegas, Nevada (2009). To *Appear*.
2. Lospinoso, J., McCulloh, I., and Carley, K. "Utility Seeking in Complex Social Systems: An Applied Longitudinal Network Study on Command and Control." Artificial Intelligence and Social Behavior Modeling Proceedings, Edinburgh, Scotland. (2009)
3. Lospinoso, J., McCulloh, I., and Carley, K. "Network Simulation Models." 26th Army Science Conference Proceedings, Orlando, Florida. (2008)
4. Lospinoso, J. "Utility Maximizing Networks: Explaining, Predicting, and Prescribing the Network." International Conference on Information & Knowledge Engineering Proceedings, Las Vegas, Nevada (2008) 344-357.
5. Baller, D., Lospinoso, J., and Johnson, A. "An Empirical Method for the Evaluation of Dynamic Network Simulation Methods." International Conference on Information & Knowledge Engineering Proceedings, Las Vegas, Nevada (2008) 358-367.
6. Baller, D., and Lospinoso, J. "Specific Communication Network Measure Distribution Estimation." 13th International Command and Control Research Technical Symposium Proceedings, Seattle, Washington. (2008)
7. Lospinoso, J. and McCulloh, I. "Social Network Probability Mechanics." 12th International Conference on Applied Mathematics of the World Science Engineering Academy and Society, Cairo, Egypt (2007), 319-325.
8. Lospinoso, J., and Moxley, F. "The ELCIT Experiment: Eliciting Organizational Effectiveness and Efficiency under Shared Belief." 12th International Command and Control Research Technical Symposium Proceedings, Washington, D.C. (2007)

Contents

Tables

Figures

INTERFACING NETWORK SIMULATIONS WITH EMPIRICAL DATA

INTRODUCTION

Current applications of Social Network Analysis (SNA) can be partitioned into two broad, non-exclusive groups: those that provide descriptions of the social network under study and those that provide prescriptions or predictions. This report explains how descriptions of the social network can advise predictions on it. Typically, SNA models are constructed based upon some theory, and sometimes their parameters are fit to empirical data. Often, however, the descriptive statistics from this estimation serve as the only prescription and prediction power from the analysis. We advocate an extension of this analysis into the realm of social network simulation to harness the full power of the empirical data available.

We survey the most popular SNA empirical models to determine which models are most effective at describing different kinds of data. We then survey popular simulation methods and underlying theories to determine how the SNA empirical models can be used to advise analysts on how to craft the simulations. Along the way, we create a suite of statistical tools which future researchers can use to determine the best analysis workflow for their particular applications.

A Background of Social Network Modeling

SNA examines relationships between social entities (e.g. people, groups, tasks, beliefs, knowledge, etc.). These entities are modeled with nodes or vertices and their connections or relationships are modeled with edges. Not all nodes are connected, and some nodes may have multiple connections. This mathematical model is applicable in content areas such as communications, information flow, and group or organizational affiliation (Tichy, 1979; Wasserman, 1994). SNA thus relies heavily on graph theory to make predictions about network structure.

Nodes are defined in terms of a set of n verticies, $V = \{v_1, v_2, \dots, v_n\}$. The nodes are related to each other with a set of edges $E = \{e_{ij}\}$, where e_{ij} is a relationship between node v_i and v_j. A social network is often shown as an adjacency matrix, where the rows and columns correspond to the nodes and each cell a_{ij} can take on any numerical value corresponding to the edge e_{ij}. In an unweighted network, cells are Boolean and are represented as 0/1: the presence or absence of an edge or relationship between nodes *i* and *j*. Networks where relationships between nodes are always mutual are called undirected networks, and their adjacency matrices will always be symmetric. Directed networks, on the other hand, can model both mutual and directional relationships. A value of 1 in cell a_{ij} represents a directed relation from node *i* to node *j*. In application, the diagonal of the adjacency matrix is rarely populated with anything but zeros, since interactions from an entity to itself are not generally interesting in a social network.

The potential complexity of interactions within even a small network, while discrete, grows exponentially with the number of entities. For this reason, algorithmic approaches to exploring state-spaces within constrained networks become computationally challenging. In a directed network, the number of possible relationships among nodes can be found by the following expression, where *n* represents the number of nodes in the network, as in $n^2 - n$.

The number of possible configurations (states) of a network with a specified number of nodes (n) and edges (ε) can be thought of as the number of unique combinations of ε nodes within the network:

$$C_{\varepsilon}^{n^2-n} = \frac{(n^2-n)!}{\varepsilon! \cdot (n^2-n-\varepsilon)!}$$

It follows that the total number of possible network configurations with n nodes can be represented by the following:

$$2^{n^2-n} - 1$$

For example, a network of 30 nodes over a dichotomous and directed relation has 7.87 x 10^{261} unique states.

To understand the probability of network structures occurring, the degree of the nodes is often investigated (Albert, 2002; McCulloh et. al., 2007; Borgotti et. al., 2006). The degree of a node, k_i, is a simple network measure counting the number of edges going into/coming out of a particular node. It is often a powerful and accurate at determination of who holds the power and influence within a network (Newman, 2007; Casciaro et. al., 1999). If we accept the notion that a random network is one in which nodes have an equal and unchanging probability to have a relationship with all other nodes in the network, random networks have a well behaved underlying distribution of degree measures. Both the degree of a node and the number of edges in a network both will follow a binomial distribution. As the network gets arbitrarily large, the distribution converges to a Poisson distribution.

There are many alternative views on what constitutes a random network; nevertheless, empirical work has shown that social networks do not construct themselves in the image of a Binomial random graph (Watts, 1998; Barabasi, 2003). Travers (1969) and Milgram (1967) studied social connections in the United States and discovered surprisingly short path lengths, where many strangers were connected by mutual acquaintances. This was termed a small-world network. A network is a small world network if its average path length is much smaller than the number of nodes in the network. This phenomenon in real-world networks is popularly known as “six degrees of separation” (Guare, 1990). Watts and Strogatz (1998) proposed the clustering coefficient as a graph level measure to indicate whether a graph is a small-world network. The clustering coefficient for a directed graph is defined as,

$$C_i = \frac{\left|\{e_{jk}\}\right|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{jk} \in E,$$

where N_i the neighborhood for a vertex v_i and is defined as its immediately connected neighbors,

$$N_i = \{v_j\} : e_{ij} \in E.$$

Intuitively, this clustering coefficient tells us how dense a nodes’ neighborhood is.
The degree k_i of a vertex is the number of vertices, $|N_i|$ in it’s neighborhood $|N_i|$. Albert and Barabasi (2002) review current methods of constructing random graphs throughout the field of Network

Science and compare the degree distribution, clustering coefficient, and average path length of multiple real-world networks with various types of random networks. They find that real-world networks have a higher average clustering coefficient and a shorter average path length than randomly generated, binomial networks with the same number of nodes and edges. Furthermore, they show that several networks have degree distributions that follow a power-law distribution, which means that very few nodes have a large degree, and many nodes have a small degree.

All of these models observe some phenomena in nature and attempt to construct some explanation for the underlying process producing the observed state. Unfortunately, it is difficult to reverse-engineer processes and validate them in this way. This paper sets out to survey some of these models (and construct a new one) and connects them with simulation packages readily available to the research community. In doing so, we bridge the gap between two disparate sections of social network analysis.

The Way Ahead

This paper proceeds by first creating a simple SNA model, the Link Probability Model, in the next chapter. The following chapter compares the LPM against popular competing models. After these SNA models are compared, we analyze how to use these models to properly parameterize Construct (a simulation package), and then test its performance. In the last chapter, we introduce an advanced topic in social network modeling that entails computationally intensive statistical methods and comment on future extensions.

THE LINK PROBABILITY MODEL (LPM)

Barabasi (2002) proposed the scale-free graph which creates a condition on the random graph that the degree distribution must follow a power law distribution. These networks were shown to resemble some real-world networks. While scale-free networks may appear to be similar to real-world networks in terms of structure, they are not a sufficient framework to truly understand the stochastic nature of networks. A new framework for random networks is proposed, based upon empirical data collected on real-world networks. This new approach produces networks that have equivalent properties to the scale-free networks; however, it is constructed in such a manner as to describe the close relationships between some nodes and distant relationships between others. This framework holds the promise of a new line of research to explore the stochastic behavior of networks.

The Link Probability Model posits that dynamic networks are constructed in the following way: considering each dyadic tie, the modeler assigns a distribution of time between communications. Integrating over this distribution according to the time between observed networks yields adjacency matrices. This generation process defines the Link Probability Model. Various methods can be used to estimate the dyadic distributions, including method of moments and maximum likelihood. Alternately, researchers can use empirical data to bootstrap dyadic distributions or simply take averages of mean time between communications. The following sections present the LPM in more formal detail.

Problem Formulation

Individuals in a social network are not connected to other individuals with uniform random probability. The probability structure is more complex. Intuitively, there are some people whom a person will communicate with or be connected more closely than others. In a study of email communication conducted at the U.S. Military Academy (McCulloh et. al., 2007), one subject emailed his wife more than ten times per day on average, while other people that he worked with received an email from him once or twice per month. For this reason, real-world networks tend to have clusters or cliques of nodes that are more closely related than others (Newman, 2003; Carley, 1996; Topper, 1999). This can be simulated by varying the probability of communication between certain nodes.

Consider a group consisting of 15 individuals, organized into three subgroups. Individuals within each subgroup work closely together and communicate more frequently than they do with people in the larger group. Each day individuals may communicate with others in the group, but most likely not everyone. If we suppose that an individual will communicate with someone in their subgroup with probability 0.8 and communicate with someone outside their subgroup with probability 0.2, we have a link probability model (LPM) shown in Figure 1.

Using this LPM, Monte Carlo simulation was used to generate 5000 instances of the network. At a 95% significance level, the confidence interval around the average clustering coefficient was 0.463 ± 0.0014 compared with 0.329 ± 0.0024 in a random graph of uniform probability. The graph generated with the LPM has a clustering coefficient that is comparable to a scale-free graph with the

same number of nodes and edges. It can be conjectured that the clustering coefficient will become greater as the within group edge probability increases. Furthermore, as the probability of certain key nodes being connected to others increases, the degree distribution will more closely follow a power law distribution. The newly proposed random network, therefore, achieves equivalent performance as the scale-free network in modeling real-world networks, yet preserves the flexibility to model dyadic relationships between nodes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A		0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
B	0.8		0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
C	0.8	0.8		0.8	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
D	0.8	0.8	0.8		0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
E	0.8	0.8	0.8	0.8		0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
F	0.2	0.2	0.2	0.2	0.2		0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2
G	0.2	0.2	0.2	0.2	0.2	0.8		0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2
H	0.2	0.2	0.2	0.2	0.2	0.8	0.8		0.8	0.8	0.2	0.2	0.2	0.2	0.2
I	0.2	0.2	0.2	0.2	0.2	0.8	0.8	0.8		0.8	0.2	0.2	0.2	0.2	0.2
J	0.2	0.2	0.2	0.2	0.2	0.8	0.8	0.8	0.8		0.2	0.2	0.2	0.2	0.2
K	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2		0.8	0.8	0.8	0.8
L	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8		0.8	0.8	0.8
M	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.8		0.8	0.8
N	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.8	0.8		0.8
O	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.8	0.8	0.8	

Figure 1. Network Probability Matrix.

The edge probabilities can be derived from empirical data in several ways. Given network data collected over multiple time periods on a group of subjects, the edge probabilities can be estimated by the proportion of edge occurrences, e_{ij}, for each cell in the adjacency matrix, a_{ij}. In the case of communication networks, statistical distributions can be fit to the time between messages for each potential edge in the network. For a specified period of time, t, the edge probability p for each set of entities i and j can be found. Let x_{ij} be the time between messages in a communication network. The probability density function for any x can then be defined as $f_{ij}(x \mid \theta_{ij})$, where θ_{ij} is the set of parameters for the distribution. Then, the probability, p, of an edge occurring within some time period t is the probability that $x < t$, which can be expressed as,

$$p = \int_0^t f_{ij}(x \mid \theta_{ij})\,dx$$

In practice, the function $f_{ij}(x \mid \theta_{ij})$ must be estimated using techniques such as maximum likelihood estimation from empirical data collected on the group being studied. It may be desirable to construct

a network based on a restriction such as, "two emails within a time period demonstrate a relationship, but one does not." In this case, it is necessary to compose a function of random variables. If $h_{ij}(2 \mid t, \theta_{ij})$ represents the probability density function of time between two sets of two emails and $f_{ij}(x \mid \theta_{ij})$ represents the probability density function of time between one set of two emails, then the following is true under certain assumptions:

$$h_{ij}(2 \mid \theta_{ij}) = \left(\int_0^t f_{ij}(x \mid \theta_{ij}) dx \right)^2$$

It is possible to generalize this idea; if $h_{ij}(x \mid \theta_{ij}, t)$ is the probability that *x* or more communications occur within time *t*, then the following is true:

$$h_{ij}(x \mid \theta_{ij}, t) = \left(\int_0^t f_{ij}(y \mid \theta_{ij}) dy \right)^x$$

This newly proposed framework for viewing the probability space of a social network preserves the same flexibility for modeling dyadic relationships, however, it provides researchers with a means to understand the probability space of the network and thus devise more robust and appropriate statistical tests for social network analysis.

Example Problem Solution

Researchers at the U.S. Military Academy monitored the e-mail traffic of 24 mid-grade (senior captains and junior majors) Army officers for 24 weeks as they were in a one year graduate program at Columbia University. Email within the group was considered, while email to outside parties was thrown out. The group had been organized with a formal leadership structure among the 24 officers. They all lived on the West Point Military Installation, and they had regular social events for the officers and their families. The degree distribution followed a power law distribution like the social networks analyzed by Barabasi and Albert (2002), and Newman (2003). The time between emails for each possible pair of nodes was calculated. There were only 65 directed pairs of nodes that had greater than 30 messages over the course of 24 weeks. Statistical distributions were fit to the time between email for the 65 pairs of nodes. All of them followed a lognormal distribution. Figure 2 shows the empirical distribution of one directed pair and four distributions fit to the data: exponential, lognormal, pareto, and zipf.

One could conjecture that the parameters of the lognormal distributions may be dependent upon various social factors, such as formal position in the network, friendship, common interest, etc. Unlike traditional social network analysis, using the LPM, an analyst can use the edge probabilities as dependent variables to study the causes of relationships, communication frequency, and ultimately network structure.

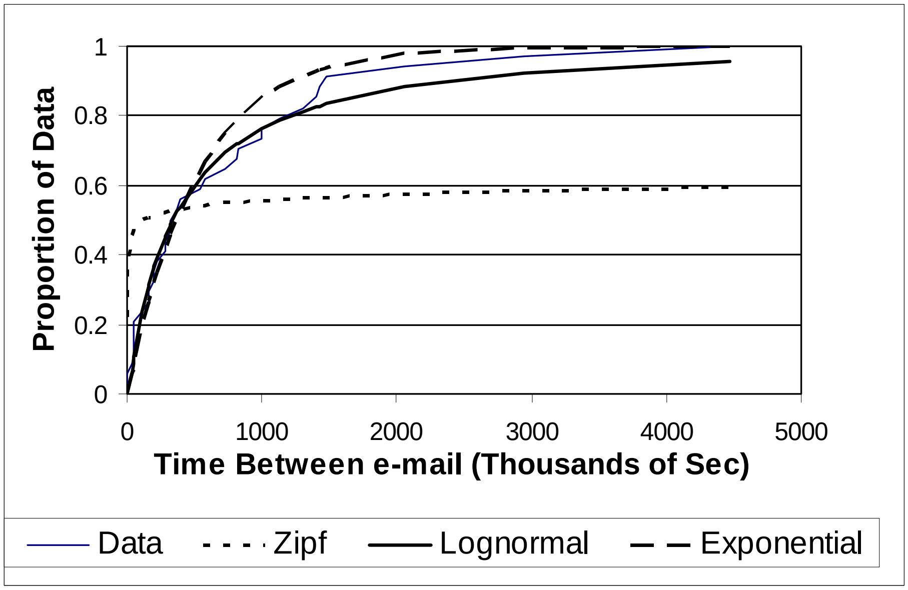


Figure 2. Distributions Fit to Time Between E-mails in Army Officer Study.

Discussion

A new approach to modeling a random network has been proposed that resembles real-world networks, preserves dyadic relationships, and can be estimated from empirical data. While the approach is surprisingly simple, it opens the door for many new analysis opportunities in social network analysis. The cell entries in the LPM can be treated as dependent variables, while various properties describing the dyadic relationships between nodal pairs can be used as independent variables. This will reduce variance in the model and increase the coefficient of determination, thereby explaining the complex behavior of a social network much better than existing methods.

Other research building from this new approach to modeling a random network can include building empirical distributions of social network measures. This newly proposed framework allows analysts to randomly generate instances of social networks under investigation. Parameters of distributions for social network measures can then be estimated using Monte Carlo simulation.

Consideration of the probability space of entity level communications is imperative for many studies of social networks. Many considerations for designing social experiments rely on conventions within the field. When constructing interaction matrices, experimenters must choose many parameters which may change the conclusion of the study. The experimenters of the U.S. Military Academy e-mail study, had to choose how many emails between two entities demonstrate a relationship to create an unweighted, directional network. To study the dynamics of the network, the experimenters further needed to determine regular intervals to sample, which allowed for a temporal analysis. By instead fitting distributions to the empirical data, experimenters could use statistical techniques to manipulate random variables and sidestep the selection of the potentially influential aforementioned parameters.

EMPIRICAL VALIDATION OF THE LPM

Presently, many structure-based frameworks are used in the network science community for the simulation of networks. These frameworks are based on the presence of triads, dyads, cliques and other network structural components. However, these frameworks do not always consider all of the factors that contribute to the dyadic relationship between agents. In a network, an agent may not be influenced by the occurrence of a triad between two other agents or that certain agents in the network have dyadic ties. The agent is mainly concerned with his own dyadic relationships—leading to an underlying dynamic equilibrium in the network.

This dynamic equilibrium is based on an underlying edge probability structure that contains a probability that each agent will communicate with every other agent in the network. The underlying probability structure of a network can remain independent of observations at any instance in time and be constant in the network under certain assumptions about its longitudinal nature and outside factors. A single observation of a tie does not necessarily designate a relationship between two agents, since the communication could have been made spuriously. On the other hand, a single observation of the lack of a tie does not designate the absence of a relationship—agents are not continuously communicating with every agent they have a relationship with at every instance in time. While a snapshot of the network at an instance in time does not indicate the dyadic relationships between agents, this snapshot is based on the underlying network probability that each agent will communicate with every other agent.

A new framework is proposed for the simulation of networks that based off of the underlying probability structure of the dynamic equilibrium. This framework is the link probability model (LPM) proposed by McCulloh, Lospinoso, and Carley (2007). The LPM estimates the edge probabilities for each dyadic pairs in the network. Probability estimation can vary from a proportion of communications in a series of observations or be estimated from more complex distributions depending on the amount and type of data present. This framework and be used to simulate a network regardless of its topology: random, small-world, scale free, cellular, etc. These LPM models require that a network is in some dynamic equilibrium, and represents the long term likelihoods that a particular dyad is observed in some state.

The edge probability structure of the underlying dynamic equilibrium remains constant in the network while the network is at a stable state. However, the underlying probabilities may change as shocks to the network take place. These probabilities may then stabilize as the network returns to it dynamic equilibrium. Using Monte Carlo simulation over an LPM will yield the underlying distributions of network measures (assuming that the network is in dynamic equilibrium). These underlying distributions can be used in change detection and allow us to statically predict shocks to the network and determine when significant changes occur, as in McCulloh (2009).

Background

Social network analysis is a theoretical framework that examines the relationships between social entities (e.g. people, groups, organizations, beliefs, knowledge, etc.). These objects are known as nodes and their connections are referred to as edges. Not all nodes are connected; however, some nodes are connected with multiple relationships. This network framework is applicable in a plethora of content areas such as communications, information flow, and group or organizational affiliation (Titchy and Tushman, 1979). Social network analysis relies heavily on graph theory to make predictions about network structure.

In 1959 mathematicians Paul Erdős and Alfréd Rénia made revolutionary discoveries in the evolution of random graphs. In their eight papers Erdős and Rénia evaluate the properties of random graphs with *n* vertices and *m* edges. For a random graph G containing no edges, at each time step a randomly chosen edge among the $\binom{n}{2}$ possible edges is added to G. This graph contains N edges and each edge of the $\binom{\binom{n}{2}}{N}$ possible edges are equiprobable. Therefore, once an edge is chosen from the $\binom{n}{2}$ equiprobable edges the next edge is chosen among the remaining $\binom{n}{2} - 1$ edges and this process is continued so that if *k* edges are fixed, all remaining $\binom{n}{2} - k$ edges have equal probabilities of being chosen (Erdős & Alfréd Rénia, 1960). A general model used to generate random graphs is as follows (Chung & Graham, 1998):

> For a given p, $0 \le p \le 1$, each potential edge of G is chosen with probability p, independent of other edges. Such a random graph is denoted by $G_{n,p}$ where each edge is determined by flipping a coin, which has probability p of coming up heads.

In this model of random graphs each edge has an equal probability of occurring or not occurring within the graph. This random graph model also assumes that all nodes in the graph are present at the beginning and the number of nodes in the network is fixed and remains the same throughout the network's life. Additionally, all nodes in this model are considered equal and are undistinguishable from each other (Barabási & Albert, 1999).

Utilizing Erdos' theory of random graphs as well as the class of uniform distributions associated with these graphs, Holland and Leinheart (1971) developed a variety of statistical tests for the analysis of social networks. Using a uniform distribution these tests spread the total probability mass equally over all possible outcomes, therefore giving an equal probability to the existence of an edge between any two nodes in the network. These statistical tests were used to develop a reference frame or constant benchmark to which observed data could be compared to determine how "structured a particular network was, or how far the network deviated from the benchmark (Wasserman and Faust, 1994)."

In 1969, Mark Granovetter proposed the strength of weak ties. In Granovetter's social world our close friends are often friends with each other as well, leading to a society of small, fully connected circle of friends who are all connected by strong ties. These small circles of

friends are connected through weak ties of acquaintances. In turn, these acquaintances have strong connections within their own circle of friends. The weak ties connecting circles of friends play an imperative role in numerous social activities from finding a job to spreading the latest fad. Close friends who have strong connections are often exposed to the same information, therefore, weak ties are activated to bridge out of our circle of friends and into the outside world (Granovetter, 1973).

Building off of Granovetter's model Duncan Watts and Steven Strogatz (1998) developed the clustering coefficient, dividing the number of links of a node's first order connections by the number of links possible between these first order connections. This clustering coefficient illustrates the interconnectivity of a circle of friends, where a value close to 1 demonstrates all first order connections of a node are connected with each other. Conversely a value close to 0 shows that a nodes first order connections are only connected through that node.

The Watts-Strogatz model of small world networks is the first to reconcile clustering with the characteristics of random graphs. According to the Watts-Strogatz model each node is directly connected to each one of its neighbors resulting in a high clustering coefficient. By clustering alone, this model has a high average path length connecting two random nodes. However, by adding only a few random links between nodes of different clusters the average separation between nodes drastically decreases. This model while containing random links between nodes keeps the clustering coefficient relatively unchanged (Watts & Newman, 1999). While the Watts-Strogatz model originally did not add extra links to the graph but randomly rewired some of the links to distant nodes the addition of random links was proposed by Watts and M. Newman.

According to Albert-László Barabási, the random graph theory of Erdős and Rénia was rarely found in the real world. Barabási has found that many real world networks have some nodes that are connected to many nodes and others that are connected to few nodes. His empirical tests showed that the distribution of the number of connections in many networks all followed a power-law distribution. These networks lack the characteristic scale in node connectivity present in random graphs, and therefore, are scale-free (Barabási, 2003). As a result of the number of connections following a power distribution, hubs are created among nodes in the network. A hub is a highly connected node that contains most of the links in the network and creates short paths between any two nodes in the network.

Barabási's model of scale-free networks is constructed around preferential attachment. For each time step a new node is added to the network. This illustrates the principal that networks are assembled one node at a time (Barabási & Albert, 1999). Assuming that each new node connects to the existing nodes of the network with two links, the probability that the new node will choose a given node is proportional to the number of links the chosen node has. Therefore, a node with more links has a higher probability of being connected to. This creates a "rich get richer" scenario where nodes with many links continue to grow by collecting new links while newer nodes with lower degrees do not collect as many links (Barabási & Albert, 1999).

Based on a scale-free network model where nodes make connections based completely on preferential attachment the probability that a new node will connect to a node with k links is

given by $\frac{k}{\sum_i k_i}$ (Barabási, 2003). This causes the first nodes in the network to develop into hub nodes due to having the longest time to collect links. However it is not always the case that the first nodes in a network develop into the biggest hubs.

To account for newer nodes overtaking older nodes as hubs, Barabási constructed the fitness model. Fitness is a nodes ability to collect links relative to every other node in the network and is based on competition in complex systems (Barabási, & Bianconi, 2001). In this new model a node's attractiveness is not determined completely by its number of links, but preferential attachment is driven by the product of the number of links a node has and its fitness. In this model the probability a new node will connect to a node with *k* links a fitness of η is $\frac{k\eta}{\sum_i k_i \eta_i}$ (Barabási, & Bianconi, 2001). Nodes in this model acquire links following the power law distribution of the scale-free model, however, the dynamic exponent η —which determines how vast a node acquires new links—is different for each node. This is proportional to a node's fitness, therefore, a node that is twice as fit as another node will obtain nodes twice as fast because its dynamic exponent is twice as large. This "fit-get-rich" model allows nodes to become hubs based on their attractiveness regardless of when they enter the network (Barabási, & Bianconi, 2001).

Contrary to the scale-free network model, Barabási, developed the "winner take all model," which strongly portrays monopolies. The "winner-take-all-model" consists of a single hub and many tiny nodes. This network develops a star topology and nodes do not acquire links following a power law distribution. McCulloh and Lospinoso (2007) proposed a new framework for random communication networks over time, based on empirical data collected on real world networks. This new framework estimates distributions for the time between communication messages, then based on a given time interval the probability of an edge occurring in the network is calculated for every ordered pair of nodes. These probabilities can be constructed through multiple techniques. To derive the probabilities from empirical data collected over several time periods, a proportion of edge occurrences can be used to estimate probabilities for each cell in the adjacency matrix.

These probabilities are displayed in a network probability matrix where each cell is the probability that node *i* communicates with node *j*. This frame work is capable of generating networks that are similar to scale free networks. Thus, this model can be used to construct any network topology: Erdős-Rénia random, Watts-Strogatz small world, Albert-Barabási scale-free, star, cellular, ect. The McCulloh-Lospinoso model is estimated from empirical data and can be used to simulate realistic observations of relationships in specific organizations.

Data

This research evaluates the density of two real world networks to find the underlying distribution of network density. The first data set was collected from a war fighting simulation in FT Leavenworth, KS in April 2007 by Craig Schreiber and Lieutenant Colonel John Graham. There were 99 participants in the experiment that were monitored over the course of four days

while data was being collected. A set of 68 participants served as staff members in the headquarters of the brigade conducting the exercise. The data displays the interactions of agents in a network which was collected by monitoring communications throughout the simulation.

The second data set is from a war fighting simulation in FT Leavenworth, KS in 2005, also collected by Craig Schreiber and Lieutenant Colonel John Graham. This data set contains 156 agents that were monitored over the course of nine iterations of the simulation. This data exhibits the communication agents in the network that was collected by monitoring communications throughout the simulation. For the duration of this chapter, the Ft. Leavenworth 2007 Data will be referred to as Network 1 and the Ft. Leavenworth 2005 data sets will be referred to as Network 2.

Method

This research explores the distribution of the density measure in two simulated networks using the network probability matrix. To simulate the network, it is necessary for a link probability model, (LPM) to be created. Once the datasets for Network 1 were trimmed of the scripted agents, they were symmetrized across the main diagonal in the Organizational Risk Analyzer (ORA) to account for the lack of directionality of communication in the data. Symmetrizing the data also corrects for the informant error of agents not reporting other agents they have communicated with. Next, the datasets from Ft. Levenworth 2007 were dichotomized to remove the weighting set by the participants. Once the data is dichotomized a one represents communication between two agents and a zero represents the lack of communication between two agents. To construct the LPM all eight data sets were compiled into a single data set consisting of the total number of discrete time periods that each agent communicated with each other agent. This matrix was then divided by the number of discrete time periods to determine the underlying edge probabilities for the network in dynamic equilibrium.

The Network 2 data sets were collected as unweighted data so they did not have to be dichotomized. It was also unnecessary to trim these data sets. The nine data sets from this network were symmetrized across the main diagonal in ORA to correct for informant of agents not reporting other agents they have communicated with. To construct the LPM all nine data sets were compiled into a single data set consisting of the total number of discrete time periods that each agent communicated with each other agent. This matrix was then divided by the number of discrete time periods to determine the underlying edge probabilities for the network in dynamic equilibrium.

The LPMs were then used as the edge probabilities for Monte Carlo simulations of these two networks. In these simulations a random number was generated for each edge. If the random number is less than the edge probability then the edge is added to the graph. This algorithm was used to create 100,000 instances of the network. When 100,000 instances of the network were completed the average density was taken from each simulation to create a dataset of 100,000 network densities for each network.

To analyze the reliability and consistency of our simulations hamming distances were utilized as a metric for the differences between two binary adjacency matrices. Using the LPM, 60,000 instances of each network were simulated. The average hamming distance from each empirical data set to every other empirical time step. Next, each simulated network is differenced in the same manner against each empirical time step. These average hamming distances were then analyzed using a paired t-test. The results of this test indicate whether the LPM predicts an instance of the empirical network with more or less error than the error introduced by the dynamic equilibriums temporal fluctuations.

The normal distribution was fit to the data of each network using Maximum Likelihood Estimation. An Anderson-Darling goodness of fit test and a comparison of the estimated cumulative distribution function to the data's empirical distribution function indicated a very good fit for the data. In addition, since the density is a linear function of the average node degree, the central limit theorem would suggest that the density is normally distributed for each network.

Using the paired t-test, it is illustrated that the networks simulated using the LPM have a smaller average hamming distance to the empirical data sets than each empirical data set is to each other. This is evidence that the simulated networks give a more reliable and consistent approximation of the underlying distribution. The results of the paired t-test for both networks are shown below in Table 1 and

Table 2 respectively. In each table column one is the average hamming distance from each empirical data set to every other empirical data set and column three is the average hamming distance from 60,000 networks simulated with the LPM to each of the empirical data sets.

Table 1. Paired t-test of Average Hamming Distances for Network 1 Data.

M	*8*	*N*	*60000*		
e_mean	e_stdev	s_mean	s_stdev	t-val	p
409.2857	38.5604	358.0939	12.77466	3.754923	0.00
365.8571	18.2978	320.0974	12.7394	7.073195	0.00
365.8571	29.04266	320.1638	12.79331	4.449958	0.00
377.8571	38.24669	330.6744	12.77289	3.489244	0.00
375.2857	36.10039	328.3765	12.79551	3.675254	0.00
349.8571	38.15944	306.0783	12.7845	3.244918	0.00
373.8571	48.45076	327.0728	12.82622	2.731135	0.01
362.4286	55.63529	317.1509	12.77754	2.301849	0.02

The p-value of each test is approximately zero indicating that there is a statistically significant difference between the empirical hamming distances and the simulated hamming distances. Additionally, since $\mu_{emperical} - \mu_{simulated} > 0$, it is shown that the simulated networks have, on average, less Hamming distance from each of the empirical data sets than the empirical data sets have from each other.

Table 2. Paired t-test of Average Hamming Distances for Network 2 Data

M	*8*	*N*	*60000*		
e_mean	e_stdev	s_mean	s_stdev	t-val	p
1445.000	84.774	1284.338	23.747	3.467	0.001
1394.750	67.487	1239.647	23.703	3.765	0.000
1296.125	85.436	1151.946	23.671	3.287	0.001
1315.875	153.533	1169.665	23.718	2.421	0.015
1191.250	112.324	1058.990	23.667	2.732	0.006
1204.875	207.944	1071.116	23.623	1.912	0.056
1167.375	190.431	1037.713	23.695	1.980	0.048
1159.625	204.465	1030.815	23.732	1.888	0.059
1170.125	195.266	1040.142	23.618	1.953	0.051

This test shows that if you select one of the empirical adjacency matrix there is more error in predicting it from the remaining empirical data sets then from predicting it with the LPM. Once the reliability and consistency of the simulations created using the LPM were confirmed, the distribution of the density could be determined. Since density is a linear function of a sample average of a network statistic and the sample sized is greater than 30 for each network the central limit theorem can be used to determine that the underlying distribution of network density is the normal distribution, with μ=0.00396148 and σ=0.0984374 for Network 1 and μ=0.0476886 and σ=0.000972361 for Network 2. This is also shown in Figure 2 and Figure 3.

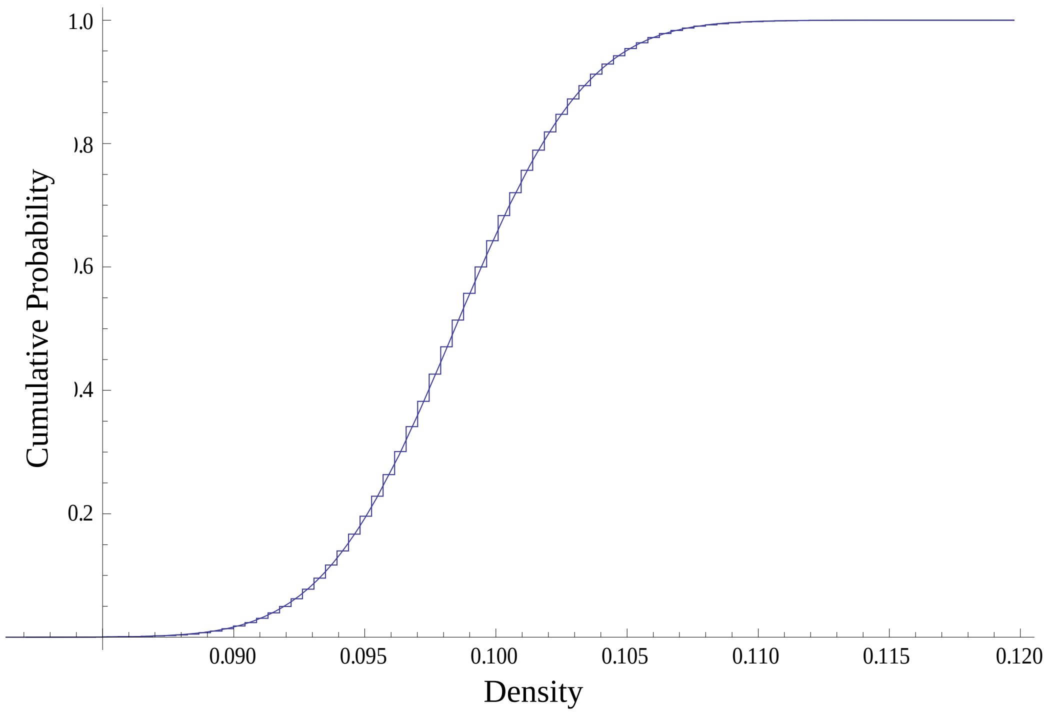


Figure 2. Stepwise Plot of Density Data for Network 1 and CDF of the Normal Distribution

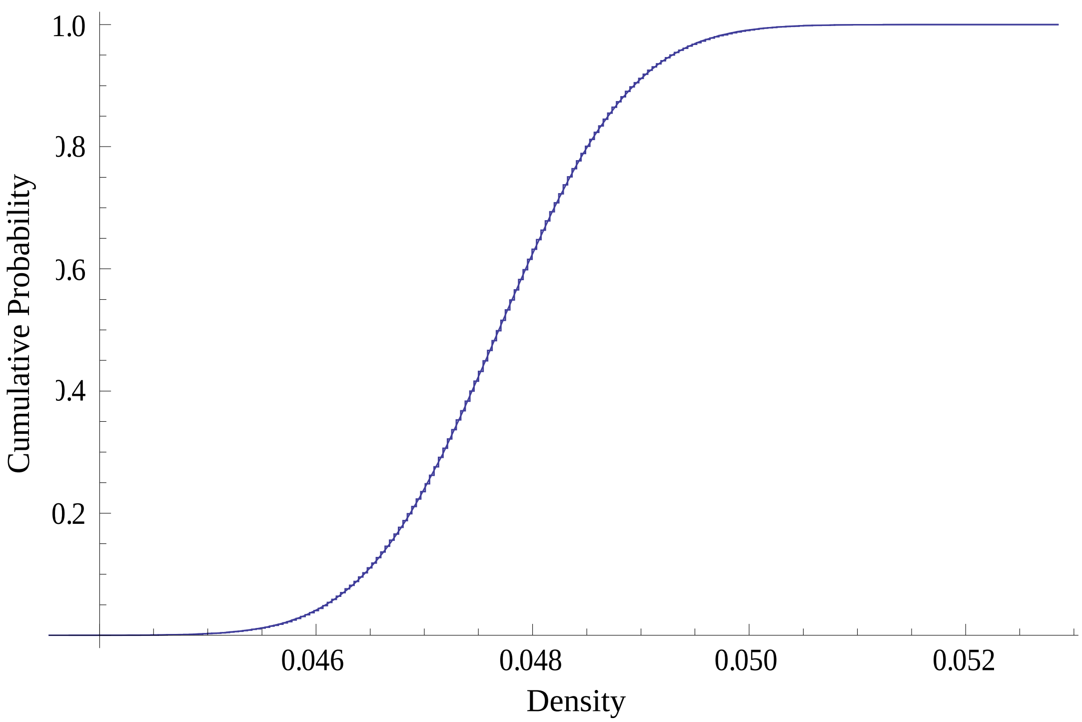
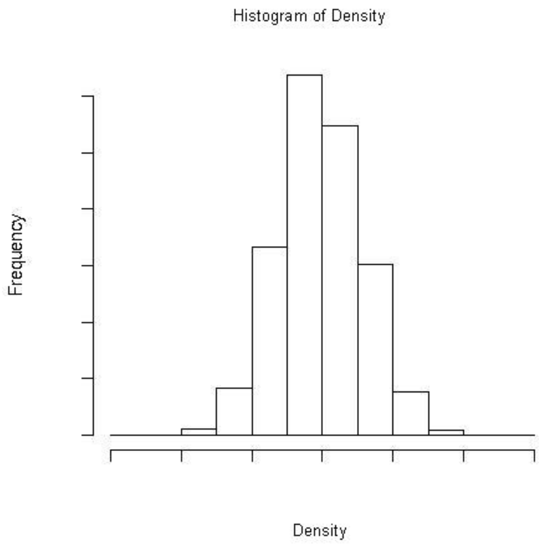


Figure 3. Stepwise Plot of Density Data for Network 2 and CDF of the Normal Distribution

Each graph shows the stepwise plot of the 100,000 densities overlaid with the CDF of the normal distribution. The sum of squared error of this model for network 1 is 9.60609 and the sum of squared error of this model for Network 2 is 1.41659. While these terms have no absolute interpretation, we can confirm upon visual inspection of Figures 2 and 3 that the data is closely fit by a normal distribution.

A histogram of the densities for Network 1 and Network 2 are shown in the figure below in Figure 4:

Histogram of Density for Network 1



Histogram of Density for Network 2

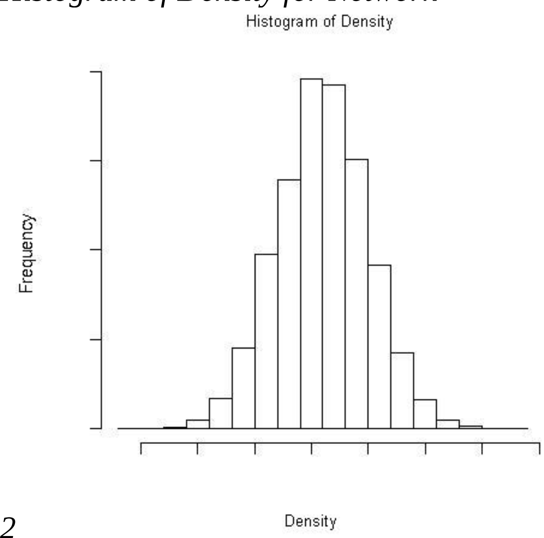
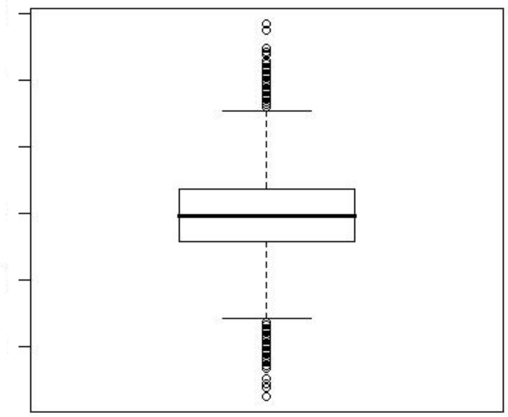


Figure 4. Histograms of Density

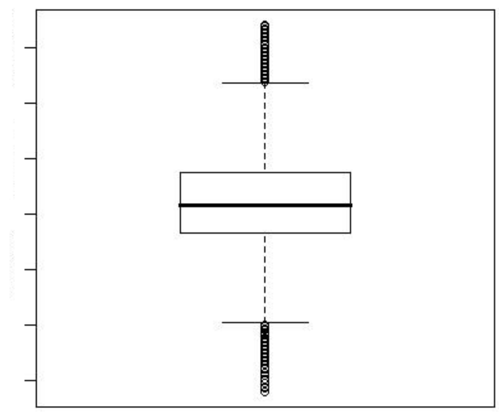
It is shown in Figure 3 that the densities of Network 1 and Network 2 both fit a normal curve. This further reinforces that the densities for both Network 1 and Normal 2 follow a normal distribution. Additional Normality tests can be seen below in Figure 5, where the box-plots indicate normal dispersion of data about the quartiles, and the qq-normal plots near-linearity indicates normality:

Box Plot of Densities for Network



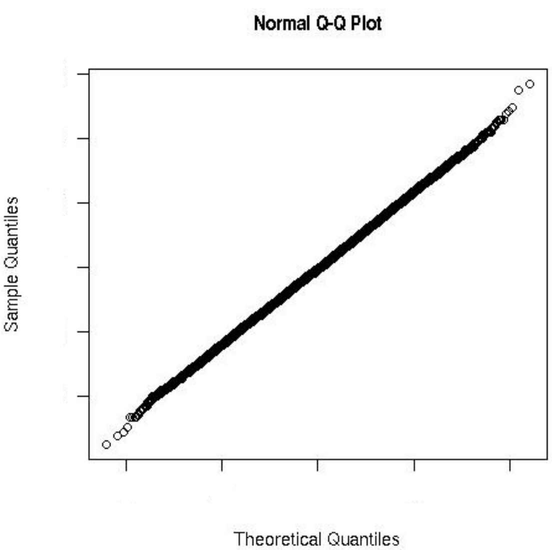
1

Box Plot of Densities for Network



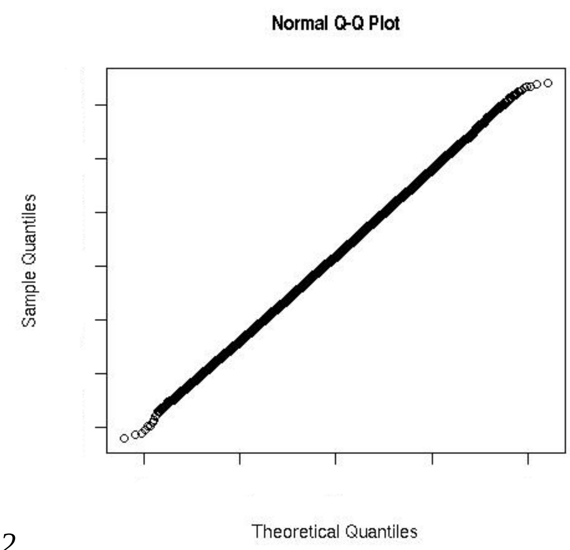
2

Normal Q-Q Plot of Densities for Network



1

Normal Q-Q Plot of Densities for Network



2

Figure 5. Additional Tests for Normality

This research validates the use of the LPM for simulating networks based on empirical data. The LPM provides a reliable and consistent network simulation that is a strong framework for analysis. This research can be extended in at least three aspects: assessing the underlying distribution for agent level statistical measures, assessing the underlying distribution for other network level statistical measures, and using these distributions to statistically predict changes and shocks to a network.

INTERFACING NETWORK SIMULATIONS WITH EMPIRICAL DATA

SIMULATING SOCIAL NETWORKS WITH CONSTRUCT

Recently, a great deal of literature has been focused on methods for simulating network structure. Simulation offers a number of advantages to the researcher. First, we can use simulation to emulate the behavior of individuals and predict behavior over time. For example, when analyzing data over time (longitudinal data), real world data at time 1 can be used to initialize the simulation program. The simulation can then be used to predict data at time 2. Second, to the extent that such predictions are accurate, we can use the simulation to do hypothetical "what if" analyses. For example, we can use the simulation program to examine alternate hypothetical societies to see what differences in such societies might be necessary to get a different outcome than that perceived in the real data.

The value of such an exercise, is not that it proves why the group or society changed as it did, but that such an exercise provides a way of reasoning about the situation, and enables the researcher to create more informed hypotheses that can then be empirically tested. In sociology, as we move to dynamic models with feedback we will find that they capture more of the social situation, but that it is incredibly difficult for the researcher to think through, without mistakes, the implications of such models. Simulation becomes a tool for increasing the specificity of theory, thinking through the theoretical implications, and generating testable predictions.

In this chapter, we provide an overview of several competing methods of network simulation. Differences and similarities are identified. The link probability model (LPM) is briefly illustrated and we identify why it is in many cases favorable to the exponential random graph (ERG) model. We then move on to summarize *Construct* and its roots in constructural sociological theory. We discover that the (LPM) provides a mathematical bridge between empirically observed data and the multi-agent simulation, *Construct*, which is based on constructuralist theory. *Construct*, in turn, introduces additional relational dependence into the LPM correcting for its naïve assumption of independence. Finally, we depict how this sociological theory translates into the LPM, how *Construct* leverages the LPM, and relate the results of empirical studies conducted by others on the effectiveness of *Construct* vice other alternatives.

Exponential Random Graph Models

ERG models are used in social network analysis as statistical models that enable an analyst to conduct inference on dependent relational data (Goodreau, 2007; Robins, et. al., 2007). The ERG model is therefore less restrictive than earlier models for social networks that assumed dyadic independence (Holland and Leinhardt, 1981). In many social network applications the relationship between two individuals depends on relationships between the individual and others in the network, cognitive limits on the number of relationships that can be maintained, similarity between individuals, and more. The ERG model framework for relaxing the dyadic independence assumption is thus essential for accurate inference in many data sets.

Estimating ERG model terms and parameters can be computationally challenging in large networks (Snijders, 2002; Pattison and Robins, 2002). Markov Chain Monte Carlo estimation of ERG models has been used to fit these models to data (Goodreau, 2007; Robins, et. al., 2007; Handcock, 2003, 2002; Snijders, 2002; Pattison and Robins, 2002). The Markov dependence in these models leads to problems of degeneracy, which is discussed in detail by Handcock (Handcock, 2003, 2002). Essentially, model degeneracy occurs when the observed data is almost impossible under the specified model. This often occurs when explanatory terms are highly correlated and there is insufficient data to construct an appropriate model. Several advances in ERG models have been proposed to include curved exponential family models (Hunter and Handcock, 2006) and neighborhood models (Robins, et. al., 2005). It is not clear that these advances have completely removed issues of model degeneracy, however.

Link Probability Model

The LPM (McCulloh & Lospinoso, 2007) has been proposed as an alternative model to the ERG model. The LPM framework for viewing the probability space of a social network avoids issues of model degeneracy, while preserving flexibility for modeling dyadic relationships. It provides researchers with an improved means to understand the probability space of the network, under certain conditions. The LPM is a square matrix where the rows and columns correspond to the nodes in a social network. The entries are the link probabilities of the directed link from the row node to the column node. This is not to be confused with an adjacency matrix, where the entries are either zero or some number representing the strength of a relationship between nodes. The link probability is a number between 0 and 1, and determines the likelihood of a link being present in an observed adjacency matrix.

The link probabilities can be derived from empirical data in several ways. Given network data collected over multiple time periods on a group of subjects, the link probabilities can be estimated by the proportion of link occurrences, $e(i,j)$, for each cell in the adjacency matrix, $a(i,j)$. In the case of communication networks, statistical distributions can be fit to the time between messages for each potential link in the network. For a specified period of time, t, the link probability p for each set of entities i and j can be found by integrating over the probability density function from 0 to t.

Relational dependence in link probabilities are accounted for in the LPM by the historic presence of links. Relational dependence in links can occur for many reasons. One example is if a boss sends an email to two employees telling them to work on a project, it will affect the probability of communication between the two employees. The LPM does not modify the link probability based on these perceived factors that may adjust the probability of two nodes having a relationship. The LPM accounts for the relational dependence, by assuming that it will be inferred by the historic presence or absence of links between nodes. If a boss often gives a task to two employees, then the presence of a link between the employees is likely to be more common when observing past networks. This does not account for all of the relational dependence in the network. To introduce a realistic degree of dependence, the LPM would need to be modified at each time step based on social theory established in the literature.

Construct for Multi-Agent Simulations

Construct[1] is a multi-agent simulation grounded in contructuralist theory (Carley, 1990; Carley, 1995). The LPM provides the stochastic engine for the multi-agent simulation. At each time step the link probabilities are determined by the nodes' perceived homophily, socio-demographics, and proximity. These social factors re-introduce the additional relational dependence missing in the raw LPM.

Construct is a dynamic-network multi-agent simulation model that can be used to examine the evolution of social, knowledge and activity networks in response to external interventions and the normal course of human interaction (Carley, 1990; Carley 1991). Network evolution and the diffusion of information and beliefs through social networks can be examined using *Construct* (Carley, 1995; Hirshman & Carley, 2007b, Hirshman, Martin & Carley, 2008). *Construct* captures group dynamic dynamics under diverse cultural and technological configurations (Schreiber & Carley, 2004). Consequently, organizational change (Carley & Hill, 2001), socio-cognitive inconsistencies (Carley & Krackhardt, 1996), the impact of communication technologies (Carley, 1995; Carley 2002) can be tested with *Construct*. To use *Construct* the researcher specifies both the agents replete with information processing capabilities (Hirshman, Carley & Kowalchuk, 2007a) and the networks in which they are embedded (Hirshman, Carley & Kowalchuk, 2007b).

Constructuralism

Before we explore the ability for network simulation to represent reality, we must first lay the foundational theory behind constructuralism as it applies to the multi-agent simulation *Construct.* Advances in both cognitive science and network theory have engendered the belief that it should be possible to develop analytical models of the relationships between individuals that would enable quantitative predictions of changes in interaction and that take into account both the self and the society, the individual and the group, the cognitive and the social. These advances have renewed interest originally seen in social comparison theory (Festinger, 1954), cognitive dissonance theory (Festinger, 1957), and balance theory (Heider, 1958), that it is possible to build a mathematics of group change as a function of individual change. It also posits that there is a gap between cognitive and individual perspectives; changes in relationships between individuals result from independent dyadic encounters. Social and structural perspective changes alter relationships between individuals. Currently a great deal of research is directed at bridging this gap. On the individual side the linking of symbolic interactionism and role theory can be viewed as a move to incorporate social or group factors into an otherwise predominantly cognitive.

Similarly, affect control theory is a move to incorporate the social, in terms of task constraints and social knowledge, into a cognitive and affective model of the individual's evaluation of; and hence determination of future action (Heise 1971, 1979, 1987; Smith-Lovin

[1] The Construct system itself is freely downloadable from the CASOS website, http://www.casos.cs.cmu.edu/projects/construct

1987). The focus on the change in the individual or his or her relationships to an actual or a generalized other, treats the group or social world as present, but relatively fixed. This implicitly assumes that social or group behavior is somehow an aggregate of the results of independent encounters between pairs of individual. This last assumption is not exclusive to those who propose more cognitively rich models of behavior.

For example, we also see it in the work on status and dominance where hierarchies are viewed to result from independent dyadic encounters (Berger, Conner, and Fisek 1974; Rosa and Mazur 1979; Lamb 1986). On the up side, evidence is being amassed that group behavior cannot be accounted for by aggregating independent dyadic encounters (Chase 1974, 1980; Ridgeway and Diekema 1989) but is rather an emergent property of the simultaneous actions of all group members (Bales 1950; Homans 1950; Chase 1974, 1980; Fararo and Skvoretz, 1986). The mechanism by which such group behavior emerges remains elusive. As a step toward locating this mechanism, research in the structural and network traditions has been moving toward providing explanations, and hence predictions, of individual cognitive change in terms of the individual's social position.

This can be seen in Burt's model of action (1982) where perceived similarity and hence norms, attitudes, likelihood of adopting innovations, and so on is a function of social position. This is further supported by Krackardt's notion (1985, 1986, 1987) that the individual's social cognition (which he defines as the individual's perception of who interacts with whom) is a function of social position. These works reveal a more cognitive actor than that revealed by classic structuralist whose behavior is nonetheless socially situated. Yet, like the more cognitive individual models, these social models of individual change, still focus on the change in the individual while maintaining a relatively fixed social world. Thus, both the individual and the social perspectives treat the social world as fundamentally stable. Consequently, neither perspective provides a mechanism by which such individual changes can produce social change. Neither approach is sufficient to explain, let alone quantitatively predict, changes in the interaction patterns for all members of the society at once. Rather, the explanations of social change are highly contextual relying on situation specific factors, forces, and constraints such as goals, coercion, bureaucratization, change in group size, and membership rituals.

Every group has a population consisting of some number of individuals. In every group, there is a set of information or facts that is potentially learnable by the members of the group. This set of information contains each piece of information that is known by at least one group member. The number of such facts will be denoted by K. The individual, for any piece of information, such as k, either knows that fact or does not. This is denoted by $F(t) = 1$ (where t denotes time) if the fact is known by individual at time period t and 0 otherwise.

Every society has a culture, which can be thought of as the distribution of information across the population. At a particular point in time, say time period t, an individual i has a certain probability to interact with another other member of the society, j. This is exactly where the LPM comes into consideration. Every society has a social structure, which can be thought of as the distribution of interaction probabilities across the population. The initial make-up of these probabilities and the transition of these probabilities at different time points are thus determined by several factors.

Construct and Constructuralism

The first assumption of the *Construct* model posits that interaction leads to shared knowledge. It is generally demonstrable that individuals acquire information (and hence will come to share knowledge) during interactions. To represent this process, a variety of simplifying assumptions are made. All pieces of information are entirely unstructured and undifferentiated. The individual may know conflicting information such as the sky is blue and the sky is green. Consequently, the overlap in what two individuals' know is just the sum of the pieces of information that they both know. When two individuals interact, each communicates one fact to the other. Individuals always learn the piece of information that is communicated to them. Consequently, if individual *i* knows that the sky is blue and individual *j* knows that the sky is green and individual *j* communicates to individual *i* that the sky is green, the overlap in their knowledge increases. Hence they have more shared knowledge. All facts known by the individual are equally likely to be communicated.

According to constructuralism, both the individual cognitive world and the socio-cultural world are continuously constructed and reconstructed as individuals concurrently go through a cycle of action, adaptation, and motivation. During this process not only does the socio-cultural environment change, but social structure and culture co-evolve in synchrony. Carley (1991a) defined the following primary assumptions in describing constructuralism: individuals are continuously engaged in acquiring and communicating information, what individuals know influences their choices of interaction partners, and an individual's behavior is a function of his or her current knowledge. In addition to these primary assumptions there were a series of implicit assumptions that upon explication serve to clarify and expand the primary assumptions. Following is an expanded list of assumptions, numbered to clarify their relation to the primary assumptions:

- 1a. Individuals, when interacting with other individuals, can communicate information.
- 1b. Individuals, when interacting with other individuals, can acquire information.
- 1c. Individuals can learn the newly acquired information, thus augmenting their store of knowledge.
- 2a. Individuals select interaction partners on the basis of relative similarity and availability.
- 2b. Individuals engage in interaction concurrently, thus an individual's first choice of interaction partner may not be available.
- 3a. Individuals have both an information processing capability and knowledge which jointly determine the individual's behavior.
- 3b. Individuals have the same information processing capabilities.
- 3c. Individuals differ in knowledge as each individual's knowledge depends on the individual's particular socio-cultural-historical background.
- 3d. Individuals can be divided into types or classes on the basis of extant knowledge differences.

These assumptions lead to a simulation template, which features a dynamic LPM as the stochastic engine. We briefly present *Construct* in this fashion, and go on to show that it performs well in simulated empirically obtained networks.

Data

The LPM and ERG models are both used to model the Sampson (1969) Monk data and the Newcomb (1961) Fraternity data, two classical datasets within the sociology literature. Sampson recorded social network data on the strength of “liking” between monks in a monastery at three different points in time. Between surveys, four of the monks were actually expelled from the monastery. The social network of these individuals was therefore changed over time. Newcomb provided 17 college transfer students with fraternity style housing in exchange for their participation in a study on friendship formation. Every week they were required to rated on a scale of 1 to 16 their preference for others in the house. Since ERG models require binary data, we use the dichotomous version of the Newcomb data proposed by Krackhardt (1998), which records a directed link between node i and node j if node i rated node j as one of their top 8 closest relationships in the network. There are 15 time periods in the Newcomb data.

Comparing the Models

The ERG model and LPM are investigated for their strengths and weakness in modeling longitudinal data in McCulloh (2008). We re-present the results here. For the Sampson (1969) monk data, an ERG model fit by Hunter, et. al. (2008) is used. An ERG model is also fit to the Newcomb (1961) fraternity data. An LPM is also fit to both the Sampson and Newcomb data sets. Monte Carlo simulation is used to generate instances of the Sampson Monk social network and the Newcomb Fraternity social network under the ERG model and the LPM.

A distance measure is required to compare the similarity between the dichotomous networks generated using the ERG model, the LPM, and the empirical data. Hamming distance (1950) is a logical choice, since it evaluates a distance between dichotomous networks. If the data were weighted networks and the models generated weighted networks as well, then a Euclidean distance would be appropriate. The quadratic assignment procedure (QAP) (Krackhardt, 1987) could be used to compare the correlation between networks; however, the correlation coefficient does not change linearly with network distance. The average Hamming distances from each empirical data set to every other empirical data set and from each simulated network to each empirical data set were calculated. These average Hamming distances were then compared using a 2-sample t-test. The results of this test indicate whether the LPM or the ERG model, models the empirical networks with more or less error.

Table 3 shows the distance between the Sampson Monk data to both the ERG and LPM.

Table 4 shows the distance between the Newcomb Fraternity data to both the ERG and LPM. It can be seen in both tables 3 and 4 that the p-values are significant at the 0.05 level. This means that there is a significant difference between how well the ERG and LPM model empirical data. The positive values for the test statistic indicate that the LPM’s average Hamming distance is less than the average Hamming distance of the empirical data. We can conclude from this test that the LPM does a significantly better job of modeling empirical data than the ERG.

Table 3. ERG and LPM Distance to Empirical Data for the Sampson Monk Data

Time period	*Mean Hamming Distance for ERG model*	*ERG Standard Deviation*	*Mean Hamming Distance for LPM*	*LPM Standard Deviation*	*T-Test Statistic*	*P-value*
1	98.7	5.697	27.67	3.5922	39.43	0.0006
2	99.1	6.2263	24.99	3.5935	37.64	0.0007
3	103.7	6.2902	24.66	3.5945	39.74	0.0006

Table 4. ERG and LPM Distance to Empirical Data for the Newcomb Fraternity Data

Time Period	*Mean Hamming Distance for ERGM*	*ERG Standard Deviation*	*Mean Hamming Distance for LPM*	*LPM Standard Deviation*	*t-test*	*p-value*
1	139.7	8.3938	91.9	5.1913	18.0147	0.0353
2	138.9	8.1847	75.1	5.2128	24.6573	0.0258
3	137.3	8.2872	48.3	5.2226	33.9732	0.0187
4	135.5	9.3363	49.7	5.2340	29.0460	0.0219
5	134.1	8.9870	50.1	5.2319	29.5558	0.0215
6	136.3	8.5251	45.5	5.2440	33.6983	0.0189
7	133.9	9.0609	47.3	5.2397	30.2202	0.0211
8	134.1	7.2946	51.9	5.2591	35.6377	0.0179
10	133.7	5.1865	64.2	5.2223	42.3990	0.0000
11	132.7	6.0562	53.4	5.2074	41.4119	0.0006
12	136.3	8.4466	51.1	5.2147	31.8930	0.0200
13	134.9	9.0117	46.6	5.2311	30.9989	0.0205
14	133.9	5.4457	46.1	5.2230	50.9574	0.0000
15	133.1	5.7242	47.2	5.2378	47.4518	0.0004

A similar test was done to compare the Hamming distance between the empirical data at each time point, with the empirical data at all other time points. The LPM was found to have no more error than that present between different time points in the empirical data. This provides evidence to validate the LPM as an effective method for simulating data.

The LPM has additional advantages. The LPM avoids the issues of model degeneracy inherent in the ERG model. The probability of link occurrence is based on the historic presence of links and does not use a Markov assumption or over specify a statistical model. For these reasons, the LPM provides an alternative method for modeling and conducting longitudinal social network analysis. For our purpose in this chapter, the LPM's ability to replicate empirical

data makes it a reasonable stochastic engine for the *Construct* multi-agent simulation model. The multi-agent simulation simply adds additional relational dependence into a model that already performs well to make it more realistic and capable of evolution over time.

Applications

The theoretical underpinnings of constructuralism as manifested in *Construct* lead us to a multi-agent simulation which utilizes a dynamic LPM as a stochastic engine for the development of knowledge diffusion and relationship building. What does this simulation provide the user?

The simulation provides an accurate, realistic simulation of social dynamics. We envision several ways in which this will be important to the military in particular and the wider academic audience in general. *Construct* can be used as a valuable decision support tool for military commanders. The social dynamics of terrorist organizations, local culture, or friendly military forces can all be modeled with the simulation. A commander can war-game potential courses of action, and evaluate alternatives using *Construct.* It can be very difficult to reason through the many potential interactions, factors, and competing theories. This simulation provides a framework that is grounded in social theory, and validated against empirical evidence, that can be used to evaluate potential courses of action.

For example, a commander might consider detaining one or more suspected terrorists. By modeling the course of action in *Construct*, he can observe the impacts of removing the individual, on the organization's performance, situational awareness, and overall effectiveness. Given limited resources, the commander could even use the simulation to optimize the individuals to remove from the social group. The simulation provides the military analyst the ability to predict the future social dynamics of an organization. This is a powerful combat multiplier for today's non-kinetic asymmetric war fighter.

The Army could also use *Construct* to evaluate the organizational structure of newly formed doctrinal units, such as the Future Combat System (FCS) operational units. The simulation can evaluate which personnel communicate more or less frequently. This can help inform efficient organization of soldiers from staff organizations to vehicle crews. Focused research on social groups can follow better experimental design, and yield greater knowledge, if an array of research questions is first evaluated in simulation. Social dynamics are complex and it can be difficult to correctly reason through different scenarios. Simulation can provide insight that may shape the research questions to be more effective.

Finally, the normal behavior of an organization can be simulated many times. From the simulations, statistical distributions can be fit to various measures of group behavior. These statistical distributions can be used to evaluate statistical hypotheses or to detect statistically significant differences between observations of the group and normal behavior. This statistical framework, therefore, increases the relevant findings one can discover in socially dynamic organizations.

We have presented two models for describing the behavior of social networks: the ERG model and the LPM. Both models were fit to two well-known data sets in the literature, the Sampson Monk data, and the Newcomb Fraternity data. The LPM modeled the data with a

statistically significant better fit than the ERG model. The benefit of the LPM was further demonstrated by finding that the difference between the LPM fit and the empirical data, was no larger than the average difference between any two samples of the empirical data.

The key limitation of the LPM is that it does not account for all of the relational dependence that is known to exist in socially connected groups. The multi-agent simulation *Construct* conveniently overcomes this limitation. *Construct* essentially uses the LPM as its stochastic engine. The link probabilities at each time step are affected by constructuralist theory established in the literature. Factors such as perceived homophily, shared knowledge, proximity, and socio-demographic variables all affect the link probabilities at each time period. These factors introduce relational dependence into the LPM. The relative weighting that these factors have can be adjusted by the user. This creates a flexible simulation tool, grounded in empirical evidence and sociological theory.

While *Construct* may be a powerful simulation tool, the current user interface limits its' capability. The Organizational Risk Analyzer (ORA) is a software package maintained by the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. ORA has an interface for near-term impact, which allows the user to isolate certain agents in a socially networked group and evaluate the impact of the isolations through simulation using *Construct.* Other than this interface, simulation runs must be conducted using an xml script. Future research will hopefully provide funding to better develop the user interface for the simulation. An improved user interface might make *Construct* available to a division ORSA to better evaluate various courses of action. This improved ability to war-game various scenarios may enhance the effectiveness of those military units.

INTERFACING NETWORK SIMULATIONS WITH EMPIRICAL DATA

ACTOR ORIENTED SOCIAL NETWORK SPECIFICATION AND ESTIMATION

Multi-agent simulation is rapidly emerging as a popular tool for understanding complex social and organizational structures. Historically, these models have been either very simple, or have contained few agents due to issues of computational complexity. As the power of computers continues to increase rapidly, more complex multi-agent simulation models are needed. Social network analysis has become equally popular for understanding social and organizational structures. This chapter applies methods in longitudinal social network analysis to multi-agent simulation.

Human organizations and social groups are composed of individuals. The individuals can be related in a number of different ways: friendship, trust, ethnicity, shared ideology, shared goals, and more. Some of these relationships are important in understanding the behavior and actions of the organization or social group. Other relationships are unimportant. Furthermore, some relationships affect others, creating very complex dynamic behavior. Multi-agent simulation is used to model individual agents that can act, interact, and learn. The agents exist in an environment where their interaction is constrained by their position in various social networks defined by the aforementioned relationships among others. Group behavior emerges as a result of the complex interaction between agents.

Understanding network structure is very important for modeling social groups and organizations in a realistic manner. For example, Valente (2007) was interested in modeling the diffusion of contraceptive innovations in the Cameroon. He found that real-world adoption rates did not follow simulation models when the network relationships were ignored. An individual’s decision to adopt an innovation is highly dependent on the decisions of adjacent individuals in a social network. Assumptions of random mixing of individuals, therefore, generate inaccurate adoption rates since trust and friendship networks are important factors. When the simulation accurately models the underlying social networks of people in the Cameroon, more accurate diffusion models are obtained. For a more thorough review of the diffusion of innovations, see Valente (2007).

Understanding social networks is not only important for modeling diffusion processes. Social networks are important for modeling any social group or organization involving humans. Multi-agent simulation modelers should be familiar with important theories in social network analysis that govern relationships between individual agents. Incorporating some of these theories into simulation models will contribute to more realistic models.

It is also important to be able to identify what social theories are applicable to certain problems and situations. Relationships that may be important in one context may be unimportant in another. Social network analysts are able to statistically test for the significance of various social theories in longitudinal network data. Equipped with significant theories governing network formation in empirical data, the multi-agent simulation modeler can include these factors in their simulation, thereby creating more realistic agent interactions.

This chapter will present a novel approach to multi-agent simulation and demonstrate it on a real-world network data set. Longitudinal network data is collected in a natural experiment focused on studying shared situational awareness and communication. An actor-oriented model (Snijders, 2007) is fit to the data to determine significant social theories contributing to network dynamics. These theories can then be incorporated in a multi-agent simulation model to create more accurate organizational behavior.

The chapter is organized as follows. First, we describe a theory of network dynamics used in social network analysis. Next, we describe the concept of network utility. In Section 4 we describe network data collected from a natural experiment conducted at the U.S. Military Academy. Section 5 describes a longitudinal analysis of that data, with the results presented in Section 6. In Section 7, we highlight implications for multi-agent simulation modelers and provide directions for future work.

Network Dynamics

Network dynamics is a term used in social network analysis to describe the behavior of networks over time (Doreian & Stokman, 1997). Social network analysts have been conducting research in this area for quite some time (Sampson, 1969; Romney 1989, Sanil et. al. 1995; Snijders, 1990; Frank 1991). There are four behaviors that can occur in a network over time: *Stability, Evolution, Random Change,* and *Mutation* (McCulloh & Lospinoso 2007; Johnson et. al., 2003).

Network *Stability* occurs when the underlying relationships that connect agents in a network remain the same over time. The observed data may contain error. Some relationships may not be observed, while some observed connections may be inadvertent and no relationship exists. Consider email communication: an agent may communicate with some friends every day, others sporadically, and they may even accidentally email someone they do not know by hitting the wrong name in a distribution list or replying to all in an email. While the observed networks may fluctuate from day to day, the underlying relationships remain unchanged. They have reached a dynamic equilibrium for at least the short term.

Network *Evolution* occurs when agent interaction over time changes the underlying relationships. Furthermore, evolution assumes that there is some underlying stochastic process that causes change over time. There are two leading approaches for modeling network evolution. One general class of approach is to use Markov chains (Wasserman, 1977, 1978, 1980). Under this approach, the network transitions from one network state to the next over time. The future state of the network is conditioned only on the current time step and not previous time steps. Research has focussed on the structure of the transition matrix that governs the evolution of the networks.

An alternate approach for modeling network evolution is multi-agent simulation (Dorean 1983; Carley 1991, 1999). Under this approach, agent based models are created in which agents interact according to some established social theory. Interactions allow the agents to change in some important way that may affect future interaction.

Random Change in a network occurs when the future behavior of the network is independent of the current state (McCulloh 2009). In other words, the agent interaction is affected by something external to the network. For example, an Army platoon may evolve as individual agents interact and communicate. When that same Army platoon comes under attack by the enemy, there is something fundamentally different about their relationships. There is not anything inherent in the individual agent interactions that could have predicted the change in network behavior as a result of the enemy attack.

It is also possible that a random change could initiate network evolution. We call this type of behavior a *Mutation.* In our Army example, it is possible that under the stress of enemy combat an individual agent displays remarkable courage or cowardice. This individual behavior may improve or remove the status of an agent. Other agents in the network may respond differently to agent based on their actions during the random change.

One possible explanation of network dynamics is agent-driven optimization. Agents in a network attempt to optimize their utility subject to various costs and constraints. Under this concept, stability can be viewed as an equilibrium surrounding some local optima. Evolution can be viewed as the network converging on some new dynamic equilibrium. Random change is still exogenous to the network and changes the state of agents in the network. If this change results in some other local optima, then the network reaches some new stability states. Otherwise, the network experiences mutation as the network converges to a new equilibrium. This concept of agent-driven optimization is further explored in this chapter as an approach for modeling complex adaptive social systems.

Network Utility

The concept of actor-driven models for network evolution was proposed by Snijders (1996). Several applications of this model have been presented. Snijders' concept of actor-driven models views a network from the perspective of individual agents. Each agent can control the set of outgoing links to other agents in the network. His seminal assumption is that actors perform myopic stochastic optimization in continuous time. These changes are Markovian and depend on network structure, attributes, and observed covariates.

Social network analysts use Snijders' actor-driven model to determine what pre-defined social factors are important in describing the evolution of empirical social network data. Snijders (2002) defines 11 basic potential objective functions that have some sociological meaning:

1. The *density effect* is defined by the number of links an agent has to other agents in the network.
2. The *reciprocity effect* is defined by the number of links to other agents that are reciprocated, in that when an agent links to a target agent, that target also links back to the original agent.
3. The *transitivity effect* is defined by the number of transitive patterns among an agent's connections. A transitive pattern occurs when two of an agent's connections are connected

themselves. This is also known as a transitive triplet. Transitivity follows the logic that two agents are more likely to know each other if they have a common friend.

4. The *balance effect* is defined by the similarity of outgoing links between an agent's connections. This theory is driven by the idea that there are positive and negative links and an agent is uncomfortable having both relations simultaneously. In other words the enemy of my friend should be my enemy and the friend of my friend should be my friend. If I am friends with my enemy's friend, I will feel uncomfortable. This effect is highly correlated with the density effect and transitivity effect. If both are included in a model a correction for the correlation between effects should be included.
5. The *number of geodesic distances of two effect* is defined by the number of other agents that an agent is indirectly connected to through an intermediary agent.
6. The *popularity effect* is defined as the number of links an agent has coming from other agents in the network.
7. The *activity effect* is defined as the number of other agents that can be reached by an agent in two steps.
8. The *main link effect* is a covariate effect for links in the network. The other objective functions might be weighted by certain relationships. For example, a link to an agent of high prestige or rank might be more valuable than a link to an agent with equivalent status.
9. The *related popularity effect* is a covariate effect for agents in the network. This is defined for an agent, *i,* as the sum of the popularity effect of all other agents connected to agent *i.*
10. The *related activity effect* is a covariate effect for agents in the network. This is defined for an agent, *i,* as the sum of the activity effect of all other agents connected to agent *i.*
11. The *related dissimilarity effect* is a covariate effect for agents in the network. This is defined as the sum of the differences in some important attribute between an agent and its' direct connections.

Agents in a network can also experience constraints as well as have objectives. Agents can be constrained in the number of links that they can maintain to other agents in the network. This constraint models cognitive limitations on individuals. A person is not capable of maintaining meaningful relationships with hundreds of people. Other constraints may be imposed on the agents in the network. Snijders does not consider constraints in his model to simplify computation. When estimating the effects, the density effect often has a negative coefficient. This is interpreted as an observed constraint on node degree. See Snijders (2002) for a more thorough explanation. Our aim is to present considerations in multi-agent simulation based on social network analysis and not to generate a comprehensive model.

Under a network utility model, an agent will change its outgoing links in such a way as to increase its overall utility, which is equivalent to optimizing its objective function (utility). It is important to note that the list of objective functions are suggestions and are non-exhaustive. When tested against empirical data, only a subset of the objective functions may be found to be significant. Undoubtedly, an analyst could consider other important social factors. Therefore, when using these objective functions in a multi-agent simulation, the modeler should use some intuition in determining important effects. Ideally, a modeler could record empirical data, use Snijders' actor-driven approach to determine significant objective function effects as his approach was intended, and then use those effects in a multi-agent simulation to make inference on the future behavior of the network.

It is important to point out differences between network utility and classic game theory. Common applications of game theory intend to focus on trading scarce resources. The network utility approach does not consider the transfer of resources, rather agents attempt to optimize their position in their social network. This approach may not be common in multi-agent simulation, but it is supported in the social sciences.

Data

Parity Communications in collaboration with the Higgins Trust Framework and the SocialPhysics project constructed the ELICIT software package. Installed on client computers, the software serves as the platform for studying organizational efficiency and effectiveness. The four phase experiment entails an introduction, practice round, a one hour exercise, and a wrap up. During both the practice round and the actual exercise, thirty four subjects are randomly assigned to one of two organizations: a typical hierarchically arrayed organization (C2) and a control-free, self-organizing organization (E). These two organizations operate independently for the duration of the exercises. See Lospinoso (2007) for more information on the experiment and basic descriptive statistics of the data.

The goal of the organization is to identify a terrorist attack based on bits of information distributed around the organization. After ten minutes of the one hour experiment, all of the correct information has been issued to the organization. Among the correct bits of information, or factoids, are also distributed false factoids. Each agent receives four factoids, and they must collaborate within the organization to come up with the correct arrangement of who, what, where, and when of the terrorist attack. The C2 group is comprised of a squad leader, four team leaders, and twelve team members. Communications among these agents are restricted to the following graph in Figure 6:

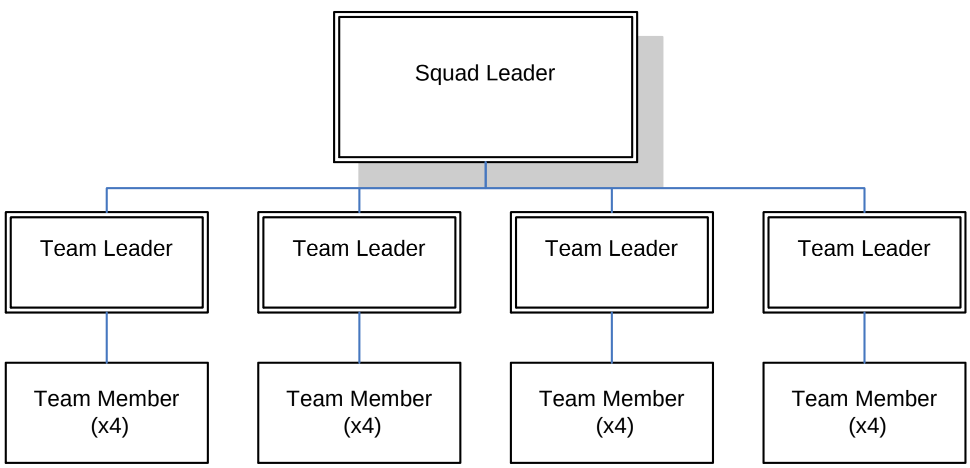


Figure 6. C2 Communications Hierarchy

Each team is dedicated to identifying one key element of the terrorist attack: who, what, where, and when.

The E group is comprised of seventeen agents with full communication capability across the organization. There are no defined teams, but the goal remains the same: positively identify the terrorist attack. All agents have the ability to post their information on their organization's website. Within the E group, this website is global to the organization. The C2 group has separate websites for each echelon (four teams and one squad site). The hierarchy in Figure 1 describes where each agent can post information. Agents can also share information with other individual agents. Once an agent believes that it knows any number of correct factoids, it can report its belief through the "identify" function to its immediate superior in the C2 group or to the entire network in the E group.

Data was collected on two iterations of ELICIT experiments conducted at West Point. During one iteration, the cadets were allowed to communicate within an edge-network configuration. In the other, the cadets were required to adhere to a strict hierarchy. Other than these systemic restrictions, the two iterations were run identically for an actual test run of two hours.

The participants in this experiment were all cadets at the U.S. Military Academy between the ages of 17 and 23. The experiment was approved for ethics and safety by the West Point Institutional Review Board. All participants received a briefing on the experiment, consented to participate, and had the option to leave the experiment at any time without any adverse impacts. The investigators conducting the experiment were not in the participants' military chain of command, so no undue influence was exerted in this experiment.

Method

We use the social network software package SIENA (Snijders et. al., 2007) which implements an actor-oriented network model to analyze data from two iterations of the ELICIT experiment. Adjacency matrices were constructed to reflect the structure of communication networks over time. These are unweighted (dichotomous), directed, and non-reflexive square matrices. We must define time intervals in which to discretize or bin the data. Following the guidelines set out by Steglich and Snijders (2007), we chose five bins. Each edge e_{ijt} was assigned a positive value (of one) if one of two conditions was met: cadet *i* sent cadet *j* information during time bin *t*, or cadet *i* posted information on a team website sometime between the start of the experiment and time *t* which cadet *j* retrieved during time *t*.

Next, we defined covariates. This step is crucial and warrants special attention when conducting an actor-oriented model specification under the SIENA framework. Covariates are empirically derived values which are infused directly into four main objective functions (effects 8-11) and provide compelling parameter estimates which can potentially gain critical insight into important aspects of sociological systems. In the case of the ELICIT data, we identify two main link effect covariates corresponding to leadership and location. The *leadership-link effect* is modeled with dependence-style network. The leadership network consists of time-invariant relationships of who was in charge of whom. Note that the leadership network was completely empty for the edge-organization case, because there were no formally defined leadership roles. The statistically significant parameter estimates of the leadership-link effect indicate that formal

leadership roles may play a significant part in driving agent behavior. With low--or even negative--parameter estimates, agents in the network are averse to forming links with formal leaders. The *location-link effect* models geographical proximity. Within the ELICIT framework, geographic distance may play a significant role within the hierarchical network, since geographical locations coincide with team placements. It would seem to also be an important covariate for the agents in the edge network, since agents within the same geographical region post to the same website and are most likely to gain information from this site. The statistically significant parameter estimates of the *location-link effect* indicate a strong affinity or aversion across both the edge and hierarchical networks on the basis of team cohesion (whether enforced or not).

In addition to main link effect covariates defined on relationships between agents, we also defined a covariate for the information an agent possesses. As time progresses in the experiment, agents gain bits of information. Once an agent believes that the information is true, they will privately publish their belief to the ELICIT server, where the belief can be recorded by the experiment administrators. This is a time varying effect. We use the related popularity effect (number 9) to model this effect. Statistically significant parameter estimates of the *information effect* indicate that agents with more information attract more communication from other agents in the network.

We also modeled the density effect, the reciprocity effect, and the transitivity effect (effects 1-3), because they are commonly used in the literature. We elected to omit other objective functions to prevent over specification of the model. See Steglich and Snijders (2006) for a more comprehensive review.

Results

To estimate the parameters of both the edge and hierarchical treatments simultaneously, we compiled both adjacency matrices and covariates into large matrices with structural holes where appropriate. We conducted estimation procedures within SIENA using default parameters and 1000 iterations of the three-stage Metropolis-Hastings Markov Chain Monte Carlso. Table 5 and Table 6 display the parameter estimates of the E and C2 networks respectively.

Table 5. Parameter Estimates for Edge Network

Measure	*Parameter Estimate (p-val)*
Density Effect	-.3693 *(.028)*
Transitivity Effect	.2054 *(.031)*
Reciprocity Effect	.1502 *(.070)*
Location-link Effect	.0513 *(.471)*
Leadership-link Effect	--
Information Effect	.2146 *(.009)*

Table 6. Parameter Estimates for Hierarchical Network

Measure	*Parameter Estimate (p-val)*
Density Effect	-.9976 *(.035)*
Transitivity Effect	.2007 *(.044)*
Reciprocity Effect	.0640 *(.36)*
Location-link Effect	.2632 *(.017)*
Leadership-link Effect	.1507 *(.023)*
Information Effect	-.1647 *(.019)*

We estimate six important objective functions to determine what sort of utility profiles are recurrent in each of the networks. After separating out the effects of each of the networks using individual covariate dummy variables, we find that the density effect measure is negative and statistically significant, which corresponds with our intuition that there is some sort of underlying cost to adding edges. Within the edge network, this effect is significantly diminished, which may indicate that agents in the edge network either have more cognitive capacity to form ties or that they are empowered by a lack of formal hierarchical structure. We find that the magnitude of this estimate (nearly -1) compared to the relative size of the other objective functions indicates that there are strong limitations to the cognitive capacity of the agents within the hierarchical network.

Transitivity effect has a strong and statistically significant, positive parameter estimate. Agents in both of these networks tend to close triads, which would confirm our intuition in the hierarchical network, where team members might be expected to close triads within their teams.[2] The estimates are rather stable across the edge/hierarchical treatment, and it would appear that there is little difference between the two utility profiles.

Reciprocity effect has little effect within the hierarchical network, but it has a significant effect on the edge network. Reciprocity tells us how likely one node is to return information to the entity who sent them information. This supports our intuition that in an edge network, relationships are created on the basis of information necessity and all agents must cross-load information. Within the hierarchical network, team-leaders can ask for information and receive information without ever having to inform their teams what is going on; so the edges are not reciprocated (which is why we fail to have statistically significant results under the hierarchical network).

Location-link effect has a statistically significant effect on the parameters for the hierarchical network. This may be a result of location and team membership being highly correlated. When two agents in the hierarchical network are within a team, their team leader tasks them with determining one of the factoids, so it is natural that collaboration here should become important. Within the edge network, there is no statistically significant estimate for location. What this indicates is that within the edge network, covariates of initial team

[2] Closing triads refers to the act of forming a relationship with a friend of a friend.

membership mean little and agents quickly breakout of their location to connect with the other locations and help contribute to their knowledge base.

Leadership-link effect was estimated for the hierarchical network and had a strong, positive estimate. This indicates that the leadership role could explain a large portion of variation in the communication patterns of the hierarchy. It both supports our intuition and supports the notion that leadership within the hierarchy was effective at promoting information sharing up and down the chain.

Information effect parameter estimates differed considerably between the edge and hierarchical treatments. Within the hierarchical network, there was actually a strong, negative correlation between people who had assembled information into some sort of conclusion and others. This means that there is information hoarding going on in the hierarchical network; the leadership is hoarding the information. Within the edge network, people who have assembled information seem to attract many edges. We cannot establish causality directly from this estimate (i.e. it could be that the entity has information because he is highly interconnected, or that he is interconnected because he has information), but it is certain that information sharing within the network is a largely significant behavioral engine.

There are some striking differences about the behavior of these two networks. First, information sharing and collaboration occurs much more within the edge network, while leadership seems to drive much of the behavior in the hierarchical network. Agents in the edge network tended develop sharing relationships much more than in the hierarchical network as evidenced by the high reciprocity and triad closure in the edge network. Finally, it appears that edge network agents had fewer constraints on collaboration *en masse* as indicated by the magnitude of their density effect estimates.

Discussion

Defense agencies of the future will increasingly rely on an understanding of complex systems. From understanding the asymmetric nature (non-hierarchical) of armed adversaries to engineering net-centric systems that maximize efficiency and effectiveness, researchers have and will continue to benefit from empirical studies of complex systems--whether social, physical, or biological. For a thorough review on this active area of research, the reader is referred to Alberts (2002).

We utilized an actor-oriented specification of a complex social system as opposed to an aggregated, holistic assessment of the system, and as a result we were able to dig into the underlying behavioral mechanics of the network and truly understand what is driving the autonomous, intelligent behavior of the cadets in the study. We now understand that soldiers within *net-centric* edge networks do collaborate across geographic and formal boundaries as expected, but more importantly--their behavior is *driven* by the need to accumulate knowledge and settle into comfortable social patterns (like triad consensus, reciprocity, etc.).

Beyond contributing to sociological literature and the defense industry's understanding of net-centric operations and systems, this chapter has introduced actor-oriented models in social network analysis which identify statistically significant utility seeking behavior within empirical data. The study of complex, adaptive systems can benefit from this empirical framework by permitting the investigator a deep look into the underlying mechanics that drive network structure. Enabled with these tools, there is a considerable array of future directions that investigators can pursue to enrich our understanding of complex systems.

Parameter estimates from an actor-oriented specification as outlined in this chapter can be used to drive a multi-agent simulation. Moreover, the approach laid out in this chapter allows a modeler to use empirical data to determine factors driving agent interaction within a simulation. Building simulation based on statistically significant findings within empirical data is an important aspect of model verification.

This approach requires that multi-agent simulation frameworks are capable of modeling significant utility seeking behavior. It is important to note that functions driving agent behavior may differ among differing applications. In the ELICIT example, different objective functions were significant for the edge and hierarchical networks, even given highly homogeneous sets of agents. This implies that there is no one model that fits all applications.
An example of a flexible multi-agent simulation is *Construct*, which is the multi-agent simulation presented earlier. In the context of Actor Oriented Models, *Construct* models agent interaction by assigning probabilities of link formation between agents at each time step. The probability of link formation is determined by a weighted function of homophily, socio-demographics, and proximity. Throughout the simulation, agents interact, share knowledge, and change in various attributes as a result of interaction with other agents. Within the framework laid out in this chapter, homophily is equivalent to transitivity, reciprocity, balance, and the information effect. Socio-demographics are equivalent to the number of geodesics of two effect, the popularity effect and the activity effect as well as some covariate effects. The proximity is equivalent to a main-link effect. Other effects can be incorporated into the *Construct* model as well. While a detailed explanation of *Construct* is beyond the scope of this chapter, we point out that it is an example of a multi-agent simulation framework that can be used to simulate empirically observed network data. The statistically significant parameter estimates of the actor-oriented model can be used to provide weights to the functions that determine the probability of link formation between agents. In this manner, the predictive power of the multi-agent simulation is enhanced due to its similarity to empirical data. Future work should explore the ramifications of resolving utility profiles into probability profiles.

An empirically grounded multi-agent simulation also contributes to better understanding network dynamics. This chapter serves to unify competing approaches to modeling network evolution. Future work may explore opportunities to introduce random change into the simulated networks. Realistic simulation of networks allows investigators to explore network dynamics by introducing various forms of evolutionary and random change at known points in time and observing their behavior. This is necessary for exploring networks over time.

The approach presented in this chapter is still limited in several ways. The list of objective function effects outlined in Section 3 is not exhaustive. There are likely other important

utility seeking functions governing agent interaction. Some effects are highly correlated and including too many effects may lead to over specified or degenerate models. Future work may investigate additional objective functions for actor-oriented models.

Multi-agent system researchers should be motivated to apply an actor-oriented approach to empirical network data. The determination of statistically significant utility seeking behavior in networks offers us a deep, complexity-preserving insight into the underlying behavior of social systems. Whether the information is used at face value to draw inference on sociological, physical, and biological phenomena, or utilized as an intermediary to simulation analysis, empirical analysis of the utility seeking behavior characterizing complex networks around us promises to deepen our understanding of them.

CONCLUSION

This paper has presented various models of social networks as longitudinally observed phenomena including the Link Probability Model, the Actor Oriented Model, and the Exponential Random Graph Model. Along the way, statistical methods were developed to differentiate among network models to determine accuracy of the models. After some analysis against empirical data from both classical literature and studies conducted at the US Military Academy, it was determined that the LPM introduces results with less difference from empirical data than the ERGM in these circumstances. We further conducted experimental studies using the ELICIT framework to test the effectiveness of the Actor Oriented Model at identifying statistically significant social theories present in the data. Finally, we found that both the LPM and the AOM fit into the social theory framework of constructuralism, which is implemented in the simulation package *Construct*.

Limitations

There are limitations on each of the modeling techniques employed. The LPM assumes dyadic independence, which is clearly not true in some circumstances of network evolution. If the network under study is in a dynamic equilibrium, however, we have found that the LPM performs well at estimating the likelihood of interactions. ERGM and AOM also have limitations in that they assume a memoryless property inherent in all Markov graph models. As we have explored in the simulation chapter, there are also some very specific assumptions made with constructuralist theory:

- 1a. Individuals, when interacting with other individuals, can communicate information.
- 1b. Individuals, when interacting with other individuals, can acquire information.
- 1c. Individuals can learn the newly acquired information thus augmenting their store of knowledge.
- 2a. Individuals select interaction partners on the basis of relative similarity and availability.
- 2b. Individuals engage in interaction concurrently thus an individual's first choice of interaction partner may not be available.
- 3a. Individuals have both an information processing capability and knowledge which jointly determine the individual's behavior.
- 3b. Individuals have the same information processing capabilities.
- 3c. Individuals differ in knowledge as each individual's knowledge depends on the individual's particular socio-cultural-historical background.
- 3d. Individuals can be divided into types or classes on the basis of extant knowledge differences.

When these assumptions do not hold, there will likely be error in the results obtained from utilizing these methods. Unfortunately, little is known on the nature of these biases.

Contributions

Contributions of this paper include preliminary analysis of the effectiveness of the ERGM, AOM, and LPM at modeling longitudinal networks, a statistical test to compare the effectiveness between these models and empirical data, and an interface for taking model parameters to teach a simulation how to represent the real world data. Much of this literature until now has existed in mutually exclusive areas of SNA. This paper serves to unify these areas and provide a framework and tools to bridge between them.

Future Work

There are many opportunities for future work. AOM must be compared against LPM and ERGM with empirical datasets. There is much work to be done in implementing an actual interface into Construct that could take empirical data and apply the SNA models with their estimation techniques directly into a simulation. In this way, researchers could obtain a body of data and seamlessly create accurate simulations of that data by specifying the appropriate multi-agent simulation model.

SNA is a rapidly expanding research area, and collaboration between social theory practitioners, statisticians, and modelers can capitalize on this expansion. This paper has illustrated how this collaboration can occur by spanning all three areas. As richer empirical data, more accurate models, and better estimation techniques become available, synthesizing them into unified suites of tools promises to deepen our understanding of networks and provide researchers with valuable and powerful insight into the social systems around us.

REFERENCES

Albert, R. and Barabasi, A. (2002). Statistical Mechanics of Complex Networks, Reviews of Modern Physics, 74: 47-97.

Alberts, D. (2002). Power to the Edge. Washington, DC: CISSP.

Bales, R.F. (1950). Interaction Process Analysis: A Method for the Study of Small Groups. Cambridge, MA: Addison-Wesley.

Barabasi, A. & Bianconi, G. (2001). "Competition and Multiscaling in Evolving Networks," in Europhysics Letters 54 (May 2001): 436-442.

Barabasi, A. (2003). Linked: How Everything is Connected to Everything Else and What It Means for Business, Science, and Everyday Life. New York: Plume.

Barabasi, A. (2003). Scale-Free Networks, Scientific American, 288:60-69.

Barabasi, A. & Albert, Réka. (1999). "Emergence of Scaling in Random Networks," in Science (286): 509-512.

Berger, J., Conner, T.L., & Fisek, M.H. (1974). Expectation States Theory: A Theoretical Research Program. Cambridge MA: Winthrop.

Borgotti, S.P., Carley, K.M., & Krackhardt, D. (2006). On the Robustness of Centrality Measures Under Conditions of Imperfect Data, Social Networks, 28:124-136.

Burt, R.S. (1982). Toward a Structural Theory of Action. New York: Academic Press.

Carley, K. M. & Hill. V. (2001). Structural change and learning within organizations. In A. Lomi and E. R. Larsen (eds), Dynamics of Organizations: Computational Modeling and Organization Theories. Menlo Park, CA: MIT Press/AAAI.

Carley, K.M. (1990). Group stability: A socio-cognitive approach. In Lawler E., Markovsky B., Ridgeway C., and Walker H. (Eds.) Advances in group processes: Theory & research. VII., 1-44. Greenwhich, CN:JAI Press.

Carley, K.M. (1991). A theory of group stability. American Sociology Review, 56(3): 331–354.

Carley, K.M. (1995). Communication Technologies and Their Effect on Cultural Homogeneity, Consensus, and the Diffusion of New Ideas. Sociological Perspectives, 38(4): 547-571.

Carley, K.M. (1996). A Comparison of Artificial and Human Organizations, Journal of Economic Behavior and Organization, 31:175-191.

Carley, K.M. (1999). On the evolution of social and organizational networks. Research in the Sociology of Organizations, 16: 3-30.

Carley, K.M. (2002). Smart agents and organizations of the future. In L. Lievrouw & S. Livingstone (eds), The Handbook of New Media, 206-220. Thousand Oaks, CA: Sage.

Carley, K.M., Krackhardt. D. (1996). Cognitive inconsistencies and non-symmetric friendship, Social Networks, 18: 1-27.

Casciaro, T., Carley, K.M., & Krackhardt, D. (1999). Positive Affectivity and Accuracy in Social Network Perception, Motivation and Emotion, 23:285-306

Chase, I.D. (1974). Models of Hierarchy Formation in Animal Societies. Behavioral Science, 19, 374-82.

Chase, I.D. (1980). Social Process and Hierarchy Formation in Small Groups. American Sociological Review, 45, 905-24.

Chung, F., & Graham, R. (1998). Erdos on Graphs: His legacy of Unsolved Problems, (Wellesley, Massachusetts: A K Peters), 74.

de Nooy, W. (2002) The dynamics of artistic prestige. Poetics, 30:147-167.

Doreian, P. & Stokman. (1997). F. Evolution of Social Networks. Gordon and Breach, Amsterdam.
Doreian, P. (1983). On the evolution of group and network structures. II. Structures within structure. Social Networks, 8: 33-64.
Erdos, P., & Rényi, A., (1960). On the Evolution of Random Graphs, Mathematical Institute of the Hungarian Academy of Science, 5:17-61.
Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999). On power-law relationships of the internet topology, Computer Communication Review, 29:251.
Fararo, T.J., Skvoretz, J. (1986). E-State Structuralism. American Sociological Review, 51, 591-602.
Faust, Katherine, & Wasserman, Stanley. (1994). Social Network Analysis: Methods and Applications, (Cambridge University Press: 1994). 506.
Festinger, L. (1954). A Theory of Social Comparison Processes. Human Relations, 7:117-140.
Festinger, L. (1957). A Theory of Cognitive Dissonance. Evanston, IL: Row, Peterson.
Frank, O. (1991). Statistical analysis of change in networks, Statistica Neerlandica 45: 283–293.
G. Van de Bunt, M.A.J. van Duijin, & Snijders, T.A.B.. (1999). Friendship networks through time: An actor-oriented statistical network model. Computational and Mathematical Organization Theory, 5: 167-192.
Goodreau, S.M., (2007). Advances in exponential random graph (p*) models applied to a large social network. Social Networks, 29, 231-248.
Granovetter, Mark S. (1973). "The Strength of Weak Ties," American Journal of Sociology 78, 1360-1380.
Guare, J. (1990). Six Degrees of Separation: A Play New York: Vintage Books.
Hamming, R.W. (1950). Error Detecting and Error Correcting Codes, Bell System Technical Journal 26(2):147-160.
Handcock, M.S., (2002). Statistical models for social networks: degeneracy and inference. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Dynamic Social Network Modeling and Analysis. National Academies Press, Washington, DC, 229-240.
Heider, F. (1958). The Psychology of Interpersonal Relations. New York: Wiley.
Heise, D. (1977). Social Action as the Control of Affect. Behavioral Science, 22: 163-77.
Heise, D. (1979). Understanding Events: Affect and the Construction of Social Action. New York: Cambridge University Press.
Heise, D. (1987). Affect Control Theory: Concepts and Models. Journal of Mathematical Sociology, 13:1-34.
Hilland, P.W., & Leinhardt, S. (1971). "Transitivity in structural models of small groups." Comparative Group Studies. 2, 107-124.
Hirshman, B.R., Carley, K.M., Kowalchuck, M.J. (2007a). Specifying Agents in Construct. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISRI-07-107.
Hirshman, B.R., Carley, K.M., Kowalchuck, M.J. (2007b.) Loading Networks in Construct. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISRI-07-116.
Hirshman, B.R., Martin, M.K., Carley, K.M. (2008). Modeling Information Access in Construct. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-08-115.

Holland, P. & Leinhardt, S. (1977). A dynamic model for social networks. Mathematical Sociology, 5:5-20.

Holland, P.W., Leinhardt, S., (1981). An exponential family of probability distributions for directed graphs (with discussion). Journal of the American Statistical Association 76, 33–65.

Homans, G.C. (1950). The Human Group. New York: Harcourt, Brace and Co.

Huisman, M. & Snijders, T.A.B. (2003). Statistical analysis of longitudinal network data with changing composition. Sociological Methods and Research, 32:253-287.

Hunter, D.R., Handcock, M.S., (2006). Inference in curved exponential family models for networks. Journal of Computational and Graphical Statistics 15, 565–583.

I. McCulloh & K.M.Carley. (2008). Detecting Change in Longitudinal Social Networks. Unpublished manuscript.

I. McCulloh & K.M.Carley. (2008). Dynamic Network Change Detection. In: Proceedings of the 26th Army Science Conference. U.S.Army, Orlando, FL.

Johnson, J., Boster, J., & Palinkas, L. (2003). Social roles and the evolution of networks in extreme and isolated environments. Mathematical Sociology, 27: 89-121.

Krackhardt, D. (1987). QAP Partialling as a Test of Spuriousness. Social Networks, 9, 171-186.

Krackhardt, D. (1998) Simmelian ties: Super strong and sticky. In Power and Influence in Organizations (eds R. Kramer, M. Neale), 21-38. Sage, Thousand Oaks, CA.

Krackhardt, D., Porter, L. (1986). The Snowball Effect: Turnover Embedded in Communication Networks. Journal of Applied Psychology, 71(1): 50-55.

Krackhardt, D., Porter, L. (1985). When Friends Leave: A Structural Analysis of the Relationship between Turnover and Stayers' Attitudes. Administrative Science Quarterly, 30(2): 242-261.

Lamb, T.A. (1986). The Familiarity Effect in Small-group Hierarchy Research. Journal of Social Psychology, 126, 51-56.

Lazarsfeld, P. & Merton, R., (1954). Friendship as social process: a substantive and methodological analysis. In Freedom and Control in Modern Society (Eds M. Berger, T. Abel, and C. Page), 18-66. New York: Van Nostrand.

Leenders, R. (1995). Models for network dynamics: a Markovian framework. Mathematical Sociology, 20: 1-21.

Lospinoso, J.A. (2007). The ELICIT Experiment: Eliciting Organizational Effectiveness and Efficiency under Shared Belief. In: Proceedings of the 12th International Command and Control Research and Technology Symposium, CCRP, Washington D.C.

Lospinoso, J.A. (2008). Utility Maximizing Networks. In: Proceedings of the 2008 International Conference on Information and Knowledge Engineering, Las Vegas, NV.

McCulloh, I., and Carley, K.M., (n.d.) The Link Probability Model: An Alternative to the Exponential Random Graph Model for Longitudinal Data. Submitted to the Journal for Social Networks.

McCulloh, I., Garcia, G., Tardieu, K., MacGibon, J., Dye, H., Moores, K., Graham, J., & Horn, D. (2007a) IkeNet: Social network analysis of e-mail traffic in the Eisenhower Leadership Development Program. (Technical Report, No. 1218). U.S. Army Research Institute for the Behavioral and Social Sciences. Arlington, VA.

McCulloh, I., Lospinoso, J. & Carley, K.M. (2007b). Social Network Probability Mechanics. In: Proceedings of the World Scientific Engineering Academy and Society 12th International Conference on Applied Mathematics, WSEAS, Cairo, Egypt.

McCulloh, Ian, & Lospinoso, Joshua. (2007) "The Probability Mechanics of Social Networks."

Milgram, S. (1967). The small world problem, Psychology Today, 2:60–67.

Newcomb, T. (1961). The Acquaintance Process. Holt, Rinehart and Winston, New York.

Newcomb, T.N. (1961). The Acquaintance Process. New York: Holt, Rinehart and Winston,

Newman, M. (2003). The Structure and Function of Complex Networks, SIAM Review, 45(2): 167-256.

Newman, M. (2007). The Mathematics of Complex Networks, Unpublished.

Pattison, P.E., Robins, G.L., (2002). Neighbourhood-based models for social networks. Sociological Methodology 32, 301–337.

Reminga, J. & Carley, K.M. (2007). ORA: Organizational Risk Analyzer v.1.7.8. Pittsburgh, PA: Institute for Software Research.

Ridgeway, C. Diekema, D. (1989). Dominance and Collective Hierarchy Formation in Male and Female Task Groups. American Sociological Review, 54, 79-93.

Robins, G.L., Pattison, P.E., (2005). Interdependencies and social processes: generalized dependence structures. In: Carrington, P., Scott, J., Wasserman, S. (Eds.), Models and Methods in Social Network Analysis. New York: Cambridge University Press, 192–214.

Robins, G.L., Pattison, P.E., Kalish, Y., Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. Social Networks, 29, 173-191.

Romney, A. (1989). Quantitative models, science and cumulative knowledge. Quantitative Anthropology, 1, 153-223.

Rosa, E. Mazur, A. (1979). Incipient Status in Groups. Social Forces, 58, 18-37.

Sampson, S. (1969). Crisis in a cloister. Ph.D. Thesis. Cornell University, Ithaca.

Sanil, A., Banks, D., & Carley, K.M. (1995). Models for evolving fixed node networks: Model fitting and model testing. Social Networks, 17.

Schreiber, C., Carley, K.M. (2004). Construct - A Multi-agent Network Model for the Co-evolution of Agents and Socio-cultural Environments. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-109.

Scott, J. (2000). Social Network Analysis: A Handbook, 2nd Ed., Newberry Park, CA: Sage.

Smith-Lovin, L. (1987). Impression Formation from Events. Journal of Mathematical Sociology, 13, 35-70.

Snijders T. A. B. & van Duijn, M.A.J. (1997). Simulation for Statistical Inference in Dynamic Network Models. In: Simulating Social Phenomena. R. Conte, R. Hegselmann, & P. Tera (Eds.) Springer, Berlin, Germany, 493-512.

Snijders, T.A.B. (1990). Testing for change in a digraph at two time points. Social Networks, 12: 539-573.

Snijders, T.A.B. (1996). Stochastic actor-oriented models for network change. Journal of Mathematical Sociology, 21, 149-172.

Snijders, T.A.B. (2007). Models for longitudinal network data. In: Models and Methods in Social Network Analysis. P. Carrington, J. Scott, Wasserman, S. (Eds.) Cambridge Press.

Snijders, T.A.B., (2001). The statistical evaluation of social network dynamics. In: Sociological Methodology. M.E. Sobel, & M.P. Becker (Eds.) Basil Blackwell, Boston, MA, 361-395.

Snijders, T.A.B., (2002). Markov chain Monte Carlo estimation of exponential random graph models. Journal of Social Structure 3, 2.

Snijders, T.A.B., C.E.G. Steglich, M. Schweinberger, & M. Huisman. (2007). Manual for SIENA version 3.1. University of Groningen: ICS / Department of Sociology; University of Oxford: Department of Statistics.

Steglich, C.E.G., Snijders, T.A.B., & West, P. (2006). Applying SIENA: An Illustrative Analysis of the Coevolution of Adolescents' Friendship Networks, Taste in Music, and Alcohol Consumption. Methodology, 2: 48-56.

Tichy, N.M., Tushman, M.L., Fombrun, C., (1979). Social Network Analysis for Organizations, The Academy of Management Review, 4(4): 507-519.

Topper, C. & Carley, K.M. (1999). A Structural Perspective on the Emergence of network Organizations, Journal of Mathematical Sociology, 24(1):67-96.

Travers, J. & Stanley M. (1969). An Experimental Study of the Small World Problem, Sociometry, 32(4): 425-443.

Valente, T. (2007). Network Models and Methods for Studying the Diffusion of Innovations. In: Models and Methods in Social Network Analysis. P. Carrington, J. Scott, Wasserman, S. (Eds.) Cambridge Press.

van Duijn, M.A.J., Zeggelink, E.P.H., Huisman, J.M., Stokman, F.N., & Wasseur, F.W. (2003). Evolution of sociology freshmen into a friendship network. Mathematical Sociology, 27: 153-191.

Wasserman, S. (1977). Stochastic Models for Directed Graphs. Ph.D. dissertation, Harvard University, Department of Statistics, Cambridge, MA.

Wasserman, S. (1979). A stochastic model for directed graphs with transition rates determined by reciprocity. In: Sociological Methodology. K.F. Schuessler (Ed.) Jossey-Bass, San Francisco CA, 392-412.

Wasserman, S. (1980). Analyzing social networks as stochastic processes. American Statistical Association, 75: 280-294.

Wasserman, S., & Faust, K., (1994). Social Network Analysis: Methods and Applications. New York: Cambridge University Press.

Watts, D.J. & Newman, M. (1999). “Renormalization Group Analysis of the Small-World Network Model,” Physics Letters, A, (263): 341

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks, Nature, 393(6684): 440-2.

Watts, J. D., & Strogatz, S. H. (1998). “Collective Dynamics of ‘Small-World’ Networks,” Nature (393). 440-442.

Weiss, L. (1953). Testing One Simple Hypothesis Against Another, Annals of Mathematical Statistics, 24:273-281.